

UNCLASSIFIED



Report Number: GA22F042

Final Report: June 1, 1999

## *Electronic Digital Imaging Standards for Archiving Records*

*By*

*Susanne H. MacTavish and Michael R. Pickard*

*Lockheed Martin Technology Services*

*Information Support Services*

*Falls Church, Virginia*

**LOCKHEED MARTIN**



Contract: GS-35F-4863G

Lockheed Martin Reference Number: GA22F042

19990604 138

DTIC QUALITY INSPECTED 4

UNCLASSIFIED

**UNCLASSIFIED**

**Approved for public release; distribution is unlimited**

(This space intentionally left blank.)

**UNCLASSIFIED**

**UNCLASSIFIED**

Imaging Standard Support Task

1 June 1999

**UNCLASSIFIED**

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>			
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 01 June 1999	3. REPORT TYPE AND DATES COVERED Final	
4. TITLE AND SUBTITLE Electronic Digital Imaging Standards for Archiving Records		5. FUNDING NUMBERS Contract: GS-35F-4863G/GA22	
6. AUTHOR(S) Susanne H. MacTavish and Michael R. Pickard			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041		8. PERFORMING ORGANIZATION REPORT NUMBER  GA22F042	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) OSD/C3I  Room 7012, Crystal Mall 3 Arlington, VA 22202		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION AVAILABILITY STATEMENT DoDD5230.24 Distribution Statement A: Approved for public release; distribution is unlimited.		12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) To improve access, distribution, and interoperability, Federal agencies are converting large numbers of documents from paper to electronic digital images. Increased accessibility to the most current data drives the move away from paper records whenever possible. Among Federal agencies there is increasing interest in receiving National Archives and Records Administration (NARA) guidance identifying acceptable Digital image formats for long term preservation. In June 1998, the Office of the Assistant Secretary of Defense Command, Control, Communications, and Intelligence (OASD/C3I) awarded a Task Order (Imaging Standard Policy Support, GS-35F-4863G/GA22) to Lockheed Martin to continue its study of digital imaging standards for archiving records. Under this Task Order the Lockheed Martin team has gathered information from the literature, interviews, and consensus gathering sessions, with a focus on three specific categories of documents that have traditionally been transferred to NARA for long-term preservation: personnel records; manuals, standards, directive type material; and documents scheduled for declassification or redact items. This report documents the current status of the three study focus areas, and provides information about digital image format options, along with associated cost and migration strategies.			
14. SUBJECT TERMS Digital Imaging Standards DoD-NARA Scanned Images Standards Conference		15. NUMBER OF PAGES 96 + x	
		16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR

Standard Form 298 (Rev. 2-89) (EG)  
Prescribed by ANSI Std. Z39.18

UNCLASSIFIED

**UNCLASSIFIED**

Imaging Standard Support Task

1 June 1999

**UNCLASSIFIED**

**UNCLASSIFIED**

Imaging Standard Support Task

1 June 1999

(This space intentionally left blank.)

**UNCLASSIFIED**

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

## Executive Summary

In 1996, the Department of Defense (DoD) initiated a study of digital image format standards. The goal was to identify and evaluate alternative electronic digital image standards for the storage and retrieval of DoD digital image records. The report of that study, Electronic Imaging Standards for Archiving Records, was issued on May 31, 1997. The report recommended that DoD pursue a strategy of adopting image standards that are embodied in commercial off-the-shelf products.

Recognizing that a significant volume of DoD records, that had been traditionally transferred to the National Archives and Records Administration (NARA) as paper records, are now being created in or converted to digital image formats, DoD asked NARA to participate in its continuing imaging standards study. Thus ensuring that NARA's long-term preservation and access needs, as well as DoD's operational record requirements, are addressed.

In June 1998, the Office of the Assistant Secretary of Defense Command, Control, Communications and Intelligence (OASD/C3I) awarded a Task Order (Imaging Standard Policy Support, GS-35F-4863G/GA22) to Lockheed Martin to continue the study, and initiate actions required for implementation of selected recommendations made in the 1997 report. Under this Task Order the Lockheed Martin team has gathered information from the literature, interviews, and consensus gathering sessions, with a focus on three specific categories of documents that have traditionally been transferred to NARA for long-term preservation:

- personnel records,
- manuals, standards, directive type material, and
- documents scheduled for declassification or redact items.

**It is clear that Electronic Records have become a very HOT topic.**

- The use of computers is changing the way government documents are created, accessed and managed. Electronic records, the Internet and E-mail have become an increasingly large part of the everyday work environment. To improve access, distribution, and interoperability, Federal agencies are converting large numbers of documents from paper to electronic digital images. Increased accessibility to the most current data drives the move away from paper records whenever possible. Among these Federal agencies there is increasing interest in receiving National Archives and Records Administration (NARA) guidance identifying acceptable digital image formats for long term preservation.
- Long-term preservation of digitally imaged records has become problematic for Federal records requiring permanent retention. While the advantages of digitally imaged documents are tremendous, due to the relatively short life cycle of digital image technology (both hardware and software), it is commonly accepted that all formats used today will eventually become obsolete.
- Computer tapes and disks deteriorate, and the hardware and software systems on which they can be read become obsolete. For an electronic record long term preservation requires that as the technology changes that the record be migrated from one format to another and then verified to ensure no loss of data. Limiting the number of image formats to monitor for technology change becomes an essential part of long-term preservation strategy. Identification of appropriate and relatively stable formats is key to success.
- While there are currently no digital image formats that are acceptable for long-term preservation, the goal is to identify formats that are likely to live longer than others in guidelines as approved data preservation formats. By selecting such standards, agencies will be able to reduce the frequency of data reformatting required to migrate data through different standards and technology and thus to minimize the cost of digital image data preservation.

**Study Conclusions:**

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

1. Access and response to Freedom of Information Act (FOIA) requests are facilitated through electronic digitalization of records.
2. No de jure standard for digital images has reached the desired maturity level for archival purposes.
3. The hardware and software technology required for the use of digital images changes rapidly.
4. Migration costs associated with archiving of digital images of textual material are unknown.
5. The anticipated high cost associated with long-term maintenance of digital image records mandates careful screening and selection of only the most valuable digital imaged records to be accessioned into the National Archives.
6. Metadata standards have been developed, but no one standard has emerged as the most universally accepted standard for electronic image records.
7. Tag Image File Format (TIFF) and Portable Document Format (PDF), both de facto standards, are the most widely used formats for text records.
8. Joint Photographic Experts Group (JPEG), a de jure standard, is the most widely used compression standard.
9. The use of proprietary standards for producing and storing images is much more common than the use of official standards.
10. Organizations will continue to use the proprietary imaging formats due to the costs involved.
11. The key roadblock to a successful digital imaging program is the high costs associated with the program and the lack of management understanding to the need for appropriate funding in the area.
12. The lack of a format standard is no longer seen as a major issue.
13. A united government voice was needed, with strong NARA leadership and a means of sharing data.

The following phased implementation approach received general acceptance at the DoD-NARA Scanned Images Standards Conference:

1. Manage the process (records management, management and policy).
2. Study, plan, gather information through cost/benefit analysis of entire life-cycle (especially document preparation, searching, and migration).
3. Pick an interim standard during step 2, which will be accepted and supported by DoD and NARA – this will enable the cost-benefit analysis to be conducted.
4. Practice migration and preservation while documents are in active use.

### Study Recommendations:

1. Image electronic digital material in the most stable formats available preferably using the latest version, but no more than two generations prior to the latest. (e.g., for TIFF image produced in January 1999 that would be TIFF version 6, 5 or 4).
  - a. Image personnel records using TIFF for archiving, TIFF or PDF formats for access. Convert all current imaged records to one standardized TIFF format.
  - b. Image declassified records using TIFF for archiving, TIFF 6 or PDF formats for access. Convert declassified versions of historically significant records to paper, microfilm or ASCII formats.
  - c. Image manuals, standards, directive type material using TIFF, ASCII and ASCII SGML or XML tagged files for archiving. Use PDF, HTML or XML formats for dissemination.
2. Plan and budget for migration of digital images every 3-5 years with of cost equivalent to 50 – 100% of the costs associated with original imaging project.

UNCLASSIFIED

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

3. Convert documents that require long-term preservation from application format to an image format.
4. Develop standard set of access metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set.
5. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.
6. Work with NARA to:
  - a. Establish criteria for selection of digital images for accessioning in the National Archives.
  - b. Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.
  - c. Establish guidelines describing metadata that must accompany digital image when submitted for archival accessioning.
  - d. Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
  - e. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
  - f. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.

**UNCLASSIFIED**

Imaging Standard Support Task

1 June 1999

(This space intentionally left blank.)

**UNCLASSIFIED**

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>iii</b>
<b>1 SUMMARY .....</b>	<b>1</b>
<b>2 INTRODUCTION.....</b>	<b>1</b>
<b>3 METHODOLOGY .....</b>	<b>2</b>
<b>4 RESULTS AND DISCUSSION .....</b>	<b>3</b>
4.1 Current Status of the Three Focus Areas.....	3
4.1.1 Personnel Records.....	3
4.1.2 Manuals, Standards, Directive Type Material.....	3
4.1.3 Declassified Documents.....	4
4.2 Electronic Record Imaging for Preservation .....	4
4.3 Data Format Standards .....	5
4.3.1 Digital Imaging File Formats .....	6
4.3.2 File Format Usage in Government Organizations.....	7
4.3.2.1 Tag Image File Format (TIFF).....	8
4.3.2.2 Portable Document Format (PDF) .....	8
4.3.2.3 Joint Photographic Experts Group (JPEG) .....	9
4.3.2.4 Standard Generalized Mark-up Language (SGML), Hypertext Mark-up Language (HTML), and eXtensible Mark-up Language (XML).....	9
4.3.2.5 Scalable Vector Graphics (SVG) .....	10
4.3.2.6 Universal Preservation Format (UPF).....	11
4.3.3 Establishing and Maintaining an Electronic Archive.....	11
4.4 Metadata .....	11
4.5 Costs .....	13
4.6 Migration Strategies .....	17
<b>5 CONCLUSIONS.....</b>	<b>19</b>
<b>6 RECOMMENDATIONS.....</b>	<b>20</b>
<b>7 REFERENCES.....</b>	<b>20</b>
<b>APPENDIX A - FORMATS.....</b>	<b>21</b>
<b>APPENDIX B - SURVEY .....</b>	<b>31</b>
<b>APPENDIX C - SURVEY RESPONDENTS.....</b>	<b>53</b>
<b>APPENDIX D - IMAGING STANDARD FOR ELECTRONIC RECORDS - ACTION PLAN.....</b>	<b>57</b>
<b>APPENDIX E - DOD-NARA CONFERENCE .....</b>	<b>59</b>
<b>APPENDIX F - CONFERENCE ATTENDEES.....</b>	<b>75</b>

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

APPENDIX G - ACRONYMS .....	87
APPENDIX H - DEFINITIONS .....	91

## TABLE OF FIGURES

FIGURE 1. TECHNOLOGY MATURITY .....	6
FIGURE 2. IMAGE PRICES .....	15
FIGURE B1 - HOW ARE DOCUMENTS ARCHIVED? .....	35
FIGURE B2 - PURPOSE OF DOCUMENT ARCHIVE .....	36
FIGURE B3 - TYPE OF STORAGE MEDIA USED .....	37
FIGURE B4 - TYPES OF ELECTRONIC FILES OR RECORDS STORED IN A DIGITAL FORMAT .....	38
FIGURE B5 - HOW RECORDS ARE CURRENTLY BEING STORED IN THE ELECTRONIC ARCHIVES .....	39
FIGURE B6 - FILE FORMATS USED TO STORE ELECTRONIC RECORDS .....	40
FIGURE B7 - PLANS FOR ELECTRONIC ARCHIVE STORAGE .....	41
FIGURE B8 - CURRENT SIZE OF ELECTRONIC ARCHIVE .....	42
FIGURE B9 - ANTICIPATED NUMBER OF RECORDS TO BE STORED IN ELECTRONIC ARCHIVES .....	43
FIGURE B10 - PLANS TO IMPLEMENT AN ARCHIVE .....	45
FIGURE B11 - PLANS FOR FUTURE IMPLEMENTATION OF AN ELECTRONIC ARCHIVE .....	45
FIGURE B12 - DIGITAL INFORMATION STORAGE IN THE FUTURE .....	46
FIGURE B13 - WOULD LIKE TO BE GIVEN DIRECTION OR GUIDELINES ON HOW TO ESTABLISH, IMPLEMENT, AND MAINTAIN AN ELECTRONIC ARCHIVE FOR DIGITAL RECORDS AND IMAGES .....	47
FIGURE B14 - SHOULD STANDARDS FOR DIGITAL RECORDS STORAGE IN AN ELECTRONIC ARCHIVE BE PROVIDED AS GUIDELINES BY NARA OR OSD? .....	47
FIGURE B15 - STANDARDS FOR HOW DIGITAL RECORDS SHOULD BE STORED IN AN ELECTRONIC ARCHIVE SHOULD BE MANDATED BY NARA OR OSD .....	48

TABLE OF TABLES

TABLE 1. COMMON METADATA STANDARDS.....	12
TABLE 2. COST INFORMATION .....	14
TABLE 3. PRODUCING DIGITAL IMAGES FROM PAPER VS. MICROFILM .....	16
TABLE 4. DIRECT COMPARISON.....	16
TABLE B1 – ANNUAL BUDGET SPENT OR ANTICIPATED .....	44

**UNCLASSIFIED**

Imaging Standard Support Task

1 June 1999

(This space intentionally left blank.)

**1**

**UNCLASSIFIED**

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

## 2 Summary

To improve access, distribution, and interoperability, Federal agencies are converting large numbers of documents from paper to electronic digital images. Increased accessibility to the most current data drives the move away from paper records whenever possible. Among Federal agencies there is increasing interest in receiving National Archives and Records Administration (NARA) guidance identifying acceptable digital image formats for long term preservation.

In June 1998, the Office of the Assistant Secretary of Defense Command, Control, Communications and Intelligence (OASD/C3I) awarded a Task Order (Imaging Standard Policy Support, GS-35F-4863G/GA22) to Lockheed Martin to continue its study of digital imaging standards for archiving records. Under this Task Order the Lockheed Martin team has gathered information from the literature, interviews, and consensus gathering sessions, with a focus on three specific categories of documents that have traditionally been transferred to NARA for long-term preservation: personnel records; manuals, standards, directive type material; and documents scheduled for declassification or redact items. This report documents the current status of the three study focus areas, and provides information about digital image format options, along with associated cost and migration strategies.

## 3 Introduction

The use of computers is changing the way government documents are created, accessed and managed. Electronic records, the Internet and E-mail have become an increasingly large part of the everyday work environment. To improve access, distribution, and interoperability, Federal agencies are converting large numbers of documents from paper to electronic digital images. Increased accessibility to the most current data drives the move away from paper records whenever possible. Among Federal agencies there is increasing interest in receiving National Archives and Records Administration (NARA) guidance identifying acceptable Digital image formats for long term preservation.

Title 44 of the United States Code (USC) and Title 36 Code of Federal Regulations (CFR) clearly identify the roles and responsibilities of federal agencies and the National Archives and Records Administration in the preservation of records of national historical interest.

Title 44 USC provides the NARA authority. It assigns the Archivist of the United States the responsibility to provide guidance and assistance to Federal officials on the management and disposition of records, to store records in centers from which agencies can retrieve them, and to take into archival facilities and Presidential libraries, for public use, records that are, in the language of Section 2107, "determined by the Archivist of the United States to have sufficient historical or other value to warrant their continued preservation by the United States Government."

As defined in Section 3301, these records are -- all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of data in them.

Title 36 Code of Federal Regulation (CFR) section 1234 sets the rules for agencies to follow regarding Electronic Records. It states that agencies must address electronic record management and that NARA should be a player in deciding how they manage their electronic records. Agencies are required to select appropriate media and systems for storing the agency's electronic records through out their life. It further states that while an agency does not need to store records in media and formats specified in 36 CFR 1228.188, it must be willing and ready to migrate the records to the currently required transfer media and formats for all permanently valuable electronic records. 36CFR1228.188 d Formats (2) Textual documents states, "Electronic textual documents shall be transferred as plain ASCII files; however, such files may contain Standard Generalized Markup Language (SGML) tags."

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

For Federal records requiring permanent retention, long-term preservation of digitally imaged records has become problematic. While the advantages of digitally imaged documents are tremendous, due to the relatively short life cycle of digital image technology (both hardware and software), it is commonly accepted that all formats used today will eventually become obsolete.

Computer tapes and disks deteriorate, and the hardware and software systems on which they can be read become obsolete. Long term preservation requires that as the technology changes an electronic record must be migrated from one format to another and then verified to ensure no loss of data. Limiting the number of image formats to monitor for technology change becomes an essential part of long-term preservation strategy. Identification of appropriate and relatively stable formats is key to success.

While there are currently no digital image formats that are acceptable for long-term preservation, the goal is to identify formats that are likely to live longer than others in guidelines as approved data preservation formats. By selecting such standards, agencies will be able to reduce the frequency of data reformatting required to migrate data through different standards and technology and thus to minimize the cost of digital image data preservation.

## 4 Methodology

The Office of the Assistant Secretary of Defense Command, Control, Communications and Intelligence (OASD/C3I) awarded a Task Order (Imaging Standard Policy Support, GS-35F-4863G/GA22) to Lockheed Martin in June 1998. The Task Order was for support in DoD's continuing study of digital image standards and the identification of the most appropriate digital imaging standard for long-term preservation of Federal documents.

With a focus on three specific categories of documents that have traditionally been transferred to NARA for long-term preservation: personnel records, manuals, standards, directive type material, and documents scheduled for declassification, the study included:

- research, gathering information from technical literature, and interviews on the status of digital imaging standards
- conduction of a survey to determine the volume, quantity, and format of electronic images and standards that each DOD activity will store and retrieve from its own libraries or transfer to the National Archives for long term preservation.
- facilitation of DOD/NARA sponsored consensus-gathering meetings and workshops; and
- publication of recommendations and findings on imagery standards for electronic records.

In October 1998, following initial research, literature review, and interviews, a survey focusing on the current usage of electronic images and archives was sent to thirty-five DOD and other Federal agencies. The survey was disseminated and returned via Email. The results were tabulated using an Access database. A copy of the survey and results are provided at the end of this report in Appendix B.

The DoD-NARA Scanned Images Standards Conference, was held March 31- April 1, 1999, at the National Archives in College Park, Md. The conference objective was to facilitate a joint Government, academic, and industry environment which would incorporate survey findings with technical knowledge and experience to determine optimum recommendations for DOD and NARA. It was attended by over 90 individuals eager for an opportunity to learn and exchange information on the current status of imaging in DoD and other Federal agencies. The Program included an overview of imaging standards including the types and extent of their use, and the status of selected imaging projects and standards associated with imaging. A summary of the conference can be found in Appendix E of this report.

In conducting the survey and facilitating the workshops, Lockheed Martin updated and expanded on the information determined in the previous DOD studies on imaging standards. Contacts with government, selected industry, and academia representatives were initiated for the survey as well as workshop attendance. OSD/C3I and NARA personnel were kept informed of study progress and findings on a regular basis via e-mail and monthly status

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

meetings, Deliverables included both a preliminary report, which was distributed to conference attendees, this final report; monthly status reports, a task order management plan; and an action plan.

## 5 Results and Discussion

### 5.1 Current Status of the Three Focus Areas

#### 5.1.1 Personnel Records

The Official Military Personnel Files (OMPF) include active duty health records, clinical records and medical treatment records. There are four major categories of personnel material:

- Service computation (enlistment, extensions, discharge)
- Professional History (education, training, promotions, security)
- Performance (performance evaluations, photographs)
- Administrative (dependent data, medical, loans, tuition assistance)

All of the Services have converted or are converting their personnel records to digital images in the TIFF 4 format, but have not utilized common indexing and system architecture. Therefore, while all these records are in TIFF format the header information has been entered differently. The Defense Personnel Records Imaging System (DPRIS) is an OSD initiative to move toward a common operating environment for electronically querying Official Military Personnel File (OMPF) records systems. DPRIS employs Web technologies to support electronic queries of the disparate OMPF systems and speed up search response times. Since the OMPF plan ultimately is to go entirely to database records, the number of TIFF records will eventually become stable.

Military Personnel Records (MPR) for discharged and deceased veterans are maintained at the National Personnel Records Center (NPRC) in St. Louis, MO. Records are usually transferred to NPRC within six months after discharge or death.

#### 5.1.2 Manuals, Standards, Directive Type Material

This category of documents have traditionally been published and distributed in paper form. DoD recently decided to stop paper publication and to make all dissemination electronically via the web, allowing ease of access to the most recent version and the ability to print on demand when paper copy is required. These records are in Microsoft Word, Hypertext Mark-up Language (HTML), Standard Generalized Mark-up Language (SGML), and Portable Document Format (PDF) formats. They will have text and embedded pictures and graphics.

The Army Logistics Support Activity (LOGSA) reports on their web page that "Paper Technical Manuals, going...going...gone..!" LOGSA is in the process of converting paper technical manuals to CD-ROM digital media. These Electronic Technical Manuals (ETMs) are, according to LOGSA, more efficient to use and will substantially reduce deployment loads. The CD-ROMs are configured by weapon system and commodity groups with the information "tagged" to link the user with corresponding information and drawings within the document. The ETMs will be distributed using the U.S. Army Publishing Agency (USAPA) system. All conversion will be completed by the end of FY 1998 with sustainment beginning in FY 1999.

The Defense Automated Printing Service encourages the use of digital files. They identify the following reasons for using digital files:

- Reduces "hidden" cost of printing
- Reduces obsolescence
- Reduces storage costs
- Reduces transportation costs
- Allows documents to be "where you want them, when you want them"

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

- Every printed copy is an original – produced at maximum resolution of the print device
- Allows the captured data to be used for more than one purpose
- Compresses document cycle time from “author” to “user”
- Streamlines business process
- Increases customer satisfaction

## 5.1.3 Declassified Documents

Executive Order 12958, signed by President William J. Clinton on April 17, 1995, mandates that all Executive Branch records of historical interest that are older than 1976 and that are classified as National Security Information be reviewed and, with the exception of nine basic exemption categories, be declassified and made available to the public by April 2000. The number of pages in this category government-wide is uncertain but is estimated to be about 2 billion.

Since many of the records must be reviewed by several agencies it was decided to use a digitized image of the record for redaction. The document declassified project has chosen the TIFF 6 format. Committing the classified documents to digital form has allowed for greater ease in exchange of the documents for review and redaction when more than one agency is required to review the document for declassification.

The Electronic Document Interchange Standard (EDIS) is a voluntary standard for electronic document interchange among Executive Branch agencies, which review electronic images of documents. The standard governs both document metadata and document images that are to be exchanged for purposes of coordinating review, as well as minimum transfer metadata. This Standard is designed solely to provide specifications for the interchange of electronic documents and related information between systems. The Standard was developed by the Declassification program Managers Council (DPMC) Automation Working Group (AWG) and The George Washington University Declassification Productivity Research Center (DPRC) for the declassification community.

Once these 2 billion records have been declassified, they will be destined for the National Archives. This process to be carried out annually from now on, as documents reach their expected time for declassification.

## 5.2 Electronic Record Imaging for Preservation

Agency records are being produced as electronic documents at an ever-increasing rate. As noted earlier in section 2.2, many government organizations are moving away from the traditional process of producing and storing paper hard copies of documents, and are moving into the realm of maintaining documents electronically.

In order to avoid the prospect of these records becoming obsolete and unreadable, there are a number of issues associated with electronic record processing that must be addressed and dealt with. The primary purpose is to ensure these electronic records can be used at any time in the future.

The file format that the electronic records are stored in has to remain readable in the future, in order to ensure these electronic records remain useable. Currently, electronic records of documents are produced in several different ways, and each method has varying levels of risk associated with it.

The most common method of archiving electronic records is to simply store the electronic file that was generated when the document was created. This application specific format could be in ASCII text format, Microsoft Word format(s), Word Perfect format(s), Microsoft PowerPoint format, or any one of hundreds of different applications currently in use by the Federal Government today. This is the most cost-effective method of storing electronic records of documents, but possesses the highest risk of the document being lost or becoming unusable.

In today's rapidly changing technology world, the application vendors need to be constantly changing their products to keep up with their competitors. The product life cycle, from the introduction of a new version of the application to the release of the next version, is typically between 1 year and 18 months. For the most part, the newer application will maintain backward compatibility with the older version, meaning that files generated with the older

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

version of the product can be viewed, edited and printed with the newer version. However, a 100% compatibility between two subsequent versions of the same product can never be guaranteed. In order to ensure that the documents can be reproduced in the newest version of the application, the document needs to be opened in the newer version, and a quality assurance check needs to be performed. Typical errors that are discovered when a new release of the product becomes available is that embedded information in the document, such as page numbering, placement of graphics, table formatting, heading numbering etc. are lost or changed. These items have to be corrected, and the document saved in the new format in order to ensure the electronic record of the document remains useable in the new format.

A preservation option for storing electronic records is to create either a vector or raster electronic image of the document. This can be accomplished by scanning the original document and saving the resultant file, or by converting the application file directly into a graphic image file.

Both of these options produce an electronic image of the original document. The method chosen will depend on the organization's preferences and budget. The primary problem associated with these methods of creating a electronic file suitable for archival and long term storage of the resultant electronic file is that a fairly large expenditure of time and resources is typically required in comparison with just storing the application file. This method of producing an electronic record is typically not accomplished for every record produced within a government organization, and due to the cost involved should not be accomplished. The government organization has to make a decision on which documents need to be preserved, how long they need to be preserved for, and how many resources need to be expended to preserve the documents. This process in itself adds to the cost of preserving an electronic record of the document.

The file format that is used store the image of the document is, or should be, a primary consideration when deciding to create an electronic image of the document. There are a number of different image formats currently available for this purpose, and each one has its advantages and disadvantages. The following section describes some of the most widely used imaging formats, and the advantages and disadvantages of each.

## 5.3 Data Format Standards

Data format standards that have received the approval or endorsement of a standards body such as the American National Standards Institute (ANSI) or the International Organization for Standardization (ISO) are referred to as de jure standards. On the other hand, data format standards that become a standard by sheer volume of usage and acceptance by users are called de facto standards. In the digital imaging arena both types of standards have advantages and disadvantages, and all carry a certain level of risk.

De jure standards take a long time to develop and must be approved by every organization that is a member of the standards organization with interests in the area covered. They are developed and maintained by a group or board of professionals. Suggested changes and updates to the standard are carefully reviewed according to a controlled process. These standards tend to be broad in scope. Frequently de jure standards are considered cumbersome and restrictive. The biggest risk associated with de jure standards is that they will never achieve user acceptance and industry penetration. Examples of de jure standards include: US MARC, JPEG, Z39.50, BIIF (Basic Image Interchange Format), and SGML.

De facto standards spring up in response to an immediate industry need. They gain in use and popularity at the dictates of the market. They are usually maintained by the group or business that originated the standard, with no community review. These standards tend to be narrower in scope and designed for one specific purpose. They penetrate the market and become a standard by virtue of the fact that that is what is available. De facto standards are a high risk choice for those looking for long term programs. De facto standards are not rigorously enforced. Several different, incompatible versions of one standard may exist at any given time. De facto standards are generally the proprietary property of one company or organization. De facto standards migrate and change very rapidly based on the needs of the information technology (IT) user community, which can result in a de facto standard becoming obsolete in a very short period of time. Since the de facto standards are proprietary property, the availability of the standard cannot be guaranteed for any great length of time. The company that holds the rights to

the de facto standard may collapse or be taken over by another organization that does not support the same de facto standard. The economy and the financial stability of the company controlling the de facto standard play a large role in how long the de facto standard will be available for use within the IT community. Examples of de facto standards include the Tagged Image File Format (TIFF) and the Portable Document Format (PDF).

A survey was sent to various Government agencies in October 1998 to determine which electronic formats were currently being used to generate digital images of documents (see Appendix B). The top six responses to this inquiry were either de facto standards, proprietary file formats, or 'unofficial' forms of approved standards (HTML is a form of SGML).

The tables in Appendix A identify and consolidate information about the most common formats for imaging. The first table contains those in the raster format with vector orientated formats identified in the second table.

### 5.3.1 Digital Imaging File Formats

The choice of the file format used to create a digital image is critical for supporting the main function of an electronic archive, that is allowing a document that is created today to be retrieved and used at any time in the future. If the file format that is used to generate the electronic image is not supported at the time the person wishes to access the document, then chances are very high that the document file will be unreadable and unusable. For this reason, the organization maintaining the electronic archive needs to make sure that the imaging file formats for the documents contained in the electronic archive are current and up to date.

Each imaging file format has a life cycle of its own, from the development and release of the format from the developer to the stage where the format is no longer supported by the commercial industry. This life cycle for imaging file formats is illustrated in the following figure. The life cycle runs from technology innovation to obsolescence. The most common imaging file formats are shown on the graph, representing the appropriate life cycle phase for each of the most commonly used file formats.

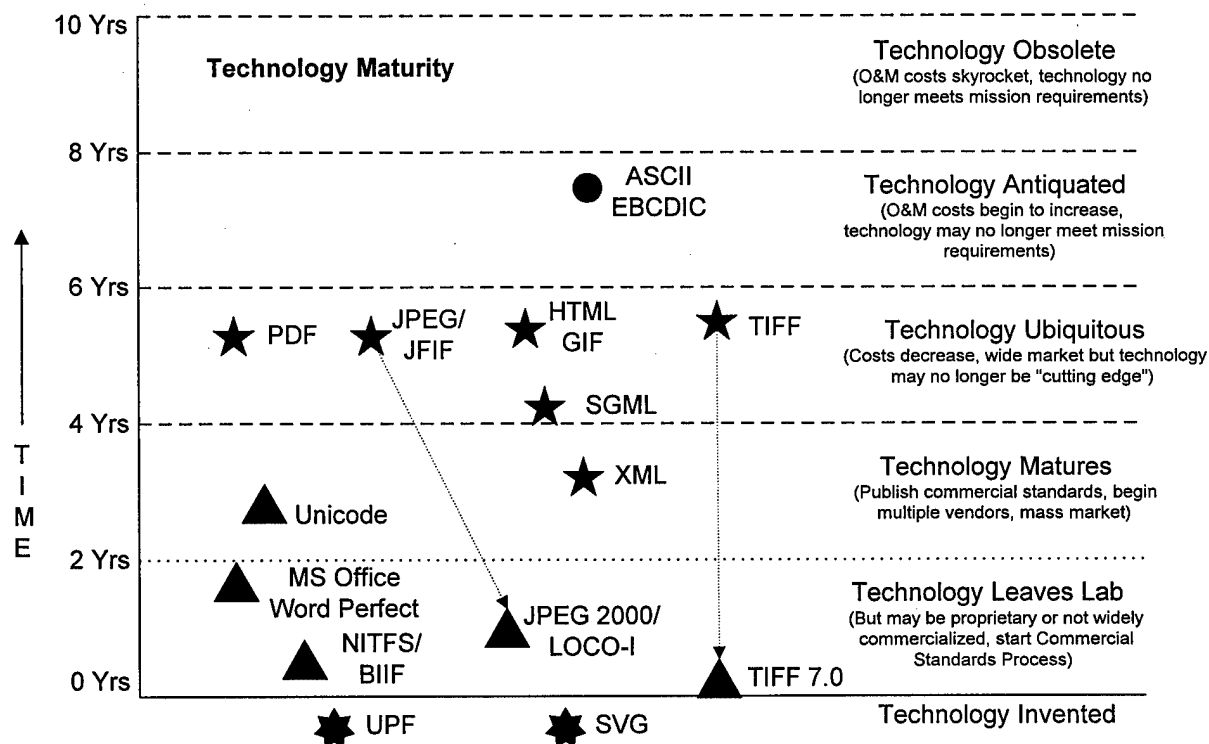


FIGURE 1. TECHNOLOGY MATURITY

This diagram also represents the maintenance costs associated with each imaging file format depending on the life cycle phase and the current migration paths for some of the formats. There are a several new file formats that are currently just over the horizon with the promise to replace the existing file formats, if there is a wide enough acceptance in the commercial marketplace. Some of the file formats listed in the figure are described in more detail in Section 4.2, based on the current popularity, acceptance and usage of the file format.

One of the most significant costs associated with the life-cycle maintenance of an electronic archive can be the migration cost of moving a document from one file format to another. The effort associated with changing the file format can be as simple as opening the document and saving it as the new format. However, experience has shown that migration is usually not this simple. When a new file format becomes available, the manufacture usually attempts to make sure that previous versions of the format are fully supported by the new format. However, there always seems to be something that doesn't work correctly, and someone has to spend a significant effort in reformatting the document file to make the new version look like the original.

An example is opening a document that was created in a previous version of Microsoft Word for Windows. Problems are usually encountered with the page layout, heading numbering, graphics, etc. even for documents that were created using the previous revision. Opening documents that were created in earlier revisions creates even more problems. This means that someone has to spend the time to reformat the document in the new format to make sure that it looks identical to the original document. If there is not a hard copy of the document available and the user does not have a working copy of the previous version of the software, then it can be extremely difficult to reformat the document to look exactly like the original.

With documents that contain only text, this much less of a problem. However, with the technology available today, word processors that make it easy to add all sorts of graphics to a document. This makes the presentation and layout of the document as important as the text itself. More and more people are becoming reliant on the old adage 'that a picture is worth a thousand words'. If the picture that represents the thousand words cannot be viewed or located when the document is opened in the future, the document itself loses a significant part of its meaning.

This is not to say that the documents created today should not contain any graphics or pictures. What this means is that the organizations that wish to create and maintain an electronic file of the document as an archived record, they need to take this into consideration when selecting the format for archiving and preserving the document.

### 5.3.2 File Format Usage in Government Organizations

In the survey of the Government organizations previously mentioned, of the organizations that currently maintain electronic document archives provided the following results when asked which file formats were currently being archived:

- 100% of the organizations archive TIFF formatted documents.
- 73% of the organizations archive PDF formatted documents.
- 55% of the organizations archive Joint Photographic Experts Group (JPEG) formatted documents.
- 45% of the organizations archive Graphics Interchange Format (GIF) formatted documents.
- 45% of the organizations archive HTML formatted documents.
- 27% of the organizations archive SGML formatted documents.
- 27% of the organizations archive Microsoft Word (.doc) formatted documents.
- 18% of the organizations archive text (.txt) formatted documents.
- 18% of the organizations archive ASCII formatted documents.
- 18% of the organizations archive Excel spreadsheet (.xls) formatted documents.
- 9% of the organizations archive PostScript formatted documents.
- 9% of the organizations archive Continuous Acquisition and Life-Cycle Support (CALS) formatted documents.
- 9% of the organizations archive Microsoft PowerPoint (.ppt) formatted documents.

## UNCLASSIFIED

- 9% of the organizations archive Word Perfect (.wpd) formatted documents.

The following sections describe the benefits and disadvantages of using the most popular file format types listed above.

### 5.3.2.1 Tag Image File Format (TIFF)

TIFF is a format standard that was developed by Aldus Corporation and Microsoft in the late 1980's as a file format designed to promote the interchange of digital image data. Since that time, there have been two major revisions to the original specification, TIFF 5.0 and TIFF 6.0. Recently, Adobe Corporation, which merged with Aldus Corporation and assumed the rights to TIFF, has announced that the TIFF 7.0 specification will be released in the near future.

As shown by the survey results, TIFF is the most common format for storing digital images of documents. The reason for this is that almost every scanner on the market today is capable of creating a TIFF file that is an exact replica of the scanned document. As shown in the survey, 91% of the organizations that maintain an electronic archive use scanning technology to create electronic images of paper documents.

One of the major advantages to TIFF is also one of its greatest liabilities when it comes to electronic archives. The TIFF file format is very flexible and loosely defined, and can be customized by the image's creator to support any number of functions such as compression, pallet colors, etc. What this really means is that not all TIFF viewers are capable of viewing all TIFF images. An example of this is that WordPerfect V5.x and V6 for IBM PC will read all base formats for TIFF, but will not read compressed TIFF files. Another indication of the problems encountered with TIFF files are the many entries in the Adobe (the maintainer of the TIFF specification) Technical Solutions Database describing problems people have encountered using the TIFF file format (**Error! Bookmark not defined.**, search on TIFF).

Adobe Corporation has not published the TIFF 7.0 specification as of the date of this document.

The costs for developing a TIFF image are minimal, and coming down steadily with the price for scanners. In the last year, the cost for a high quality scanner has dropped significantly, allowing almost every organization the luxury of owning a scanner. The scanners typically come bundled with imaging software, allowing for even more options in creating a digital image.

### 5.3.2.2 Portable Document Format (PDF)

Adobe Corporation's PDF format has become a de facto standard for publishing documents on the World Wide Web (WWW). The PDF format has several key benefits, the most significant of which is that it is a completely device independent page description language. This open systems approach has made for a wide acceptance of PDF as the standard for publishing documents either on the web or for printed documents.

The PDF file is typically much smaller than the original document format, therefore enabling more documents to be stored on the same media. Depending on the fonts used in the original document, the PDF format may produce an exact replica of the original document, including graphics, pictures, and tables.

One of the biggest disadvantages to PDF is that it is a proprietary format that is owned by the Adobe Corporation. However, Adobe has made the PDF standards available to other vendors, and other software manufacturers have created products that produce and read PDF files. The following list identifies several vendors that market PDF creation products:

- ZEON Corporation's DocuMaker program will convert any document that is saved in a postscript format to a PDF document.
- FastIO Systems provides the ClibPDF program, an ANSI C Source Library for direct PDF generation without relying on any Adobe Acrobat tools and related products.
- 5D is a company that provides the NIKNAK software tool that converts postscript files to PDF files.

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

- Adobe Corporation also provides a freeware program, PDFMaker for Microsoft Word 97 that works with Adobe Acrobat 3.0 for Windows to convert Microsoft Word documents into PDF files.

While there are a number of companies that provide product support for PDF, the format is still proprietary. This makes users dependent on Adobe if they want the latest product line that supports the generation of PDF file.

### 5.3.2.3 Joint Photographic Experts Group (JPEG)

A number of organizations reported on the survey that they were archiving documents in the JPEG format. While there is a .jpg file name extension for files using the JPEG compression method, there is not an actual file format called JPEG. There are actually at least three different file formats that use the .jpg file name extension:

- Still Picture Image File Format (SPIFF) is the official ISO standard JPEG file format.
- JPEG File Interchange Format (JFIF) is the de facto standard for JPEG images developed by C-Cube Microsystems, because it took the ISO took over five years to develop the SPIFF standard.
- Image JPEG (IMJ) was created by Pegasus Image Corporation as a variation of the JFIF file format. IMJ is essentially a JFIF file with a Microsoft Windows Bitmap (BMP) header and enhanced palette optimization. The IMJ format is used in several screensaver applications and by organizations such as Delrina and the National Center for Missing Children.

These three file formats are for the most part compatible and most JPEG readers will read all three formats. However, this is not always the case. Some JPEG readers will only open JFIF images, while still others generate an error message when attempting to open a JPEG image other than SPIFF. This problem relates to the JPEG Standard itself, which has 44 different modes for compressing images. Most of these modes are application specific.

The ISO is in the process of developing a new image compression standard called JPEG 2000. This new standard is being developed to compliment, not to replace the current JPEG standard (ISO 10918-1, ISO 10918-2, ISO 10918-3). One of the goals of the new standard is to develop a single decompression architecture that will encompass all of the different compression modes.

The baseline JPEG is classified as a lossy compression algorithm because the decompressed output is not bit-for-bit identical to the original input. The baseline JPEG compression ratio can be set to provide an output image that is visually indistinguishable from the original, but there will always be some loss of image quality. The JPEG Standard ISO 10918-3 currently contains a lossless compression algorithm, and another lossless algorithm, JPEG-LS, is in the final draft international standard FDIS14495-1.

The compressed JPEG images, since they are considered to be lossy, should not be used as the archived version of a document or image. The document that is maintained in the archive should be either the original document, or the document in a file format that is identical to the original document.

### 5.3.2.4 Standard Generalized Mark-up Language (SGML), Hypertext Mark-up Language (HTML), and eXtensible Mark-up Language (XML)

SGML, HTML and XML are all markup languages that were designed for the transmission of information from one computer to another. The differences between the three are quite distinct, but the basic format for the format files themselves remains the same.

All markup language files can be viewed using a standard text editor. The codes that are placed in the SGML, HTML or XML files that describe the formatting characteristics for the document are simple text codes placed in brackets. The actual information that is contained in the format file is stored as text data. Images are inserted into the file as hyperlinks: the actual files for the pictures, images and graphs displayed with the text data are not actually stored inside the markup language file. The hyperlinks provide the data path to the image or picture that needs to be inserted on the page.

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

On February 11, 1999 the World Wide Web Consortium (W3C) released the first working draft of the Scalable Vector Graphics (SVG) specification. This specification will allow vector graphics to be inserted as text information directly into the mark-up language file. There are several significant advantages to this new specification, the most critical is that this specification will eliminate the necessity for having more than one file. Another critical advantage to the SVG specification is that text searches can be performed on the text information contained in the vector graphic. Currently, separate metadata information about the contents of the vector graphics file must be provided if the user needs to perform a search on the image.

SGML is defined in ISO Standard 8879:1986, and is a formal language used to pass information about the component parts of a document from one computer system to another. The markups provided by SGML tell the computer that is displaying the document more than just how the information is to be displayed on the monitor, such as where it is displayed on the screen, which fonts to use, and where graphics should be inserted in the text. SGML provides a method for describing the relationships between different parts of a document, such as paragraph numbering, table of contents, indexes, etc. SGML also allows users to include metadata about the document such as the author's name, date published, etc. within the SGML file.

SGML is currently the archival imaging format of choice for many libraries because it allows users to perform a search on the text contained in the SGML file. The SGML file itself can be read and searched using a standard text editor, unlike the other imaging formats which require special software to be used such as optical character recognition (OCR) software.

HTML is an application of the SGML that uses a predefined set of document type definitions (DTDs) that are used to markup documents, describing how the document should be formatted for the user's screen. The difference between HTML and SGML is that SGML does not provide a standard set of DTDs. The document's creator can define the DTDs, and passed along with the SGML file to the computer that requests the file over the Internet.

XML is a subset of the SGML standard that is becoming more and more popular, and may one-day replace both HTML and PDF as the most prevalent web publishing formats. The difference between HTML and XML is that XML allows users to specify their own customized tags the same as with SGML, but is not possible with HTML. This capability, of letting the document writer prepare and provide their own DTD, creates an extra file that the web browser has to download from the source site to determine the meaning of the customized tags in the document.

The advantages that are available with SGML and XML that are not available with HTML are in the area of metadata. With SGML and XML, author of the documents is able to insert metadata such as the author's name, the date published, and the topic or subject of the document. The metadata can be marked with custom tags, such as <author/> or <subject/> that allows users to search for the document using this criteria. With HTML, this type of information needs to be included in with the text itself, rather than as metadata or 'data about the data.'

### 5.3.2.5 Scalable Vector Graphics (SVG)

The first working draft of the SVG specification was released by the World Wide Web Consortium (W3C) on 11 February 1999. This file format specification promises to change the way that vector graphics are inserted into the Markup Language file formats.

With the current Markup Language file formats, each graphic is contained in its own separate file. The HTML, SGML or XML document must contain a hyperlink to the graphics file. This results in the document creator having to maintain, update, and edit several different files for each document. This also means that when an electronic document is placed in the electronic archive, all graphics and text files must be present. The hyperlinks contained in the markup language document must be updated to point to the correct location, which might change every time the document is moved from one physical location or media type to another. This also prevents the document user from being able to perform searches on the text contained in the graphic file.

The SVG specification changes all of this. With SVG, the vector graphic is inserted directly into the Markup Language document eliminating the need for the hyperlink to a separate file. A second major advantage to the SVG

specification is that the text contained in the vector graphic now becomes a part of the main document, allowing user's to perform searches on the text.

Most of the major graphics software vendors, including Adobe, Apple, Autodesk, Corel, HP, IBM, Inso, Macromedia, Microsoft, Netscape, Quark, RAL, Sun, and Visio have been supporting the development of the SVG specification. This indicates that there will be wide industry acceptance for this new graphics format, and promises to change the way HTML, SGML and XML documents are generated and archived.

#### 5.3.2.6 Universal Preservation Format (UPF)

In 1996, the National Historical Publications and Records Commission of the National Archives awarded a grant to WGBH, a public broadcasting station in Boston Massachusetts, to research and produce a prototype of a platform-independent Universal Preservation Format (UPF). This file format would be designed specifically for digital technologies that will ensure the accessibility of a wide array of data types into the indefinite future. A draft document describing this initiative can be found on the WWW at **Error! Bookmark not defined.**

#### **5.3.3 Establishing and Maintaining an Electronic Archive**

There are a number of methods currently in use today for the generation of digital images of documents to be stored in an electronic archive. The choices taken by the organization that is responsible for the preservation of the document are dependent on a number of factors, cost usually being one of the most critical criteria.

The cost of creating and maintaining an electronic archive is much greater than just the cost of creating the digital image, storing the resulting electronic file on some type of media, and placing the media in a safe location. The long-term costs such as migration of records from one format or media to another must be taken into account or the organization runs the risk of not being able to retrieve documents from the archive at a later date.

The consideration in creating an electronic archive is determining which documents generated within the organization need to be preserved, and for how long these documents need to be preserved. As shown in the survey results, this will vary from organization to organization. While a majority of the organizations reported that less than 10% of the documents stored in an electronic archive need to be preserved for a long period of time, several organizations reported that up to 100% of the records stored in their electronic archive need to be permanently preserved at the National Archives. (see Appendix B)

Being able to access records stored in an electronic archive entails much more than just storing the electronic files on a network drive, CD-ROM, or optical drive. The records themselves need to be cataloged, indexed, and linked to a text file that provides an explanation of the contents of the image. If this is not accomplished, then while the records themselves may be preserved, the information contained within will not be very useful to others trying to access the records in the future.

This information about the image or electronic record is referred to as metadata.

To be truly efficient, an electronic archive should be built on a database concept, where the image can be linked to the metadata text file and other information supporting the image.

There are three factors which are critical for ensuring an electronic archive is established that minimizes the life cycle costs and ensures that the information contained in the archive can be retrieved at any time it is required:

- The format that is used to create the digital image
- The media that is used to store the digital images, and
- The use of a Records Management System (RMS) or a Document Management System (DMS), which is DoD 5015.2 compliant, to manage the electronic archive.

## **5.4 Metadata**

Metadata is data about data. In this case data or information about the image or electronic record. Metadata typically support a specific function: discovery or access; administrative; or structural. Access metadata include

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

location, subject, authors, creator, etc. Administrative metadata include type of item, file format, compression format, dimensions, bit-depth, color lookup table, etc. Structural takes administrative data one step further and identifies file size relationship to other file records.

The standard for bibliographic data is US MARC. In its complete format it is designed to be a transfer format of bibliographic data from one system to another. Many feel that MARC records are too expensive and time consuming. However, it is not necessary to do a complete AACR2 (Anglo-American Cataloguing Rules, Second Edition) catalog record to have a MARC record. You merely need to identify your selected fields or tags with the associated MARC field identifier and your record will be accessible on thousands of Commercial-Off-The-Shelf (COTS) products designed to search or access data.

The desire for increased access to electronic records and to the Web has driven initiatives such as the Encoded Archival Describing (EAD), Dublin Core, the Text Encoding Initiative (TEI) and the Resource Definition Framework (RDF) Extensible Markup Language (XML). Every electronic file format must contain some form of metadata to tell the computer how to display the record. This is usually called the header of the record and thus there is a TIFF metadata standard and a BIIF metadata standard, etc. Specialized collections of data have also created metadata standards. For example the Federal Geospatial Data Committee (FGDC) has established a metadata set for geo-spatial data (digital maps and related items). There is also the Warwick Framework, an architecture that allows for the interchange of distinct metadata packages, Z39.50, and the set of required metadata found in DoD 5015.2 Std.

Each format has structure and administrative data in its header information. Ideally this data should be standardized. However, front-end search can work through a defined set of differences. Example of this is the front-end work to provide access to Official Military Personnel Files (OMPF) records.

The table that follows is a sampling of fields from some of the most common metadata standards. It illustrates that both a core of data can be found and that specialized fields are required for different categories of images.

TABLE 1. COMMON METADATA STANDARDS

DoD 5015.2 - STD	Dublin Core	MARC	GILS	EAD
Subject	Subject/keywords	650, 653\$a	Uncontrolled term	
	Title	245\$a	Title	<titleproper>, <unittitle>
Author or originator	Author/Creator	700, 710, 720\$a	originator	<origination>
Originating organization	Publisher	260\$b	Distributor	<publisher>
	Other Contributor	720\$a, 700, 710	Contributor	
Document creation date	Date	260\$c	Date of publication	<date>, <unitdate>
Media type	Resource Type	655\$a	medium	
format	Format	856\$q	Available linkage type	<physdesc>
	Resource Identifier	856\$u	Available linkage	
	Relation	787\$n	Cross reference	

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

DoD 5015.2 - STD	Dublin Core	MARC	GILS	EAD
			relationship	
	Source	786\$n	Sources of Data	
	Language	546\$a , 041\$a	Language of Resource	
	Coverage	500\$a, 255\$c, 513\$b	Supplement information Bounding coordinates Time period textual	
	Rights	540\$a	Use constraints	
	description	520\$a	abstract	<abstract>
Date filed		X		
Addressee				
Location of record		X		
Vital record indicator				

## 5.5 Costs

The cost of preparing electronic files or records for archiving documents is a critical factor in the decision making process. However, due to a number of factors, the costs associated with the long-term preservation of electronic documents are extremely complex and very fluid. The difficulty associated with creating a cost model for archiving documents has been addressed by a number of researchers and professionals working in the electronic archiving field, but a complete cost analysis has never, to the authors knowledge, been published. At best, any cost model can only be a snapshot of the archiving costs for any given time.

The following quote illustrating this point is taken from the Cornell University Digital to Microfilm Conversion: A Demonstration Project 1994-1996, Final Report to the National Endowment for the Humanities by Anne R. Kenney:

"Numerous conferences and reports have been dedicated to issues associated with digital archiving-ensuring continuing access to digital materials across hardware/software configurations and subsequent generations of computer technology. The clearest articulation of these issues is provided in the Joint Task Force Report of the Research Libraries Group and the Commission on Preservation and Access, entitled *Preserving Digital Information: Final Report and Recommendations*. As the report makes clear, currently there are no agreed-upon processes or model institutional programs for preserving digital collections over time. **There is even less consensus on the costs of such efforts.**"

Some of the reasons for the complexity of the cost model are as follows:

- The information technology field is the most rapidly changing industry today. Almost as soon as a new product or standard becomes available on the market, it becomes obsolete due to new advances in technology, new products that hit the market and new requirements that are being identified due to these new technologies and products.
- The costs for information technology are constantly fluctuating, not only decreasing in some areas but increasing in others as well.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

There are a number of factors that must be looked at when determining the total costs associated with archiving electronic documents. The costs for creating the electronic file or record are not the only costs that must be looked at. The person or group preparing the electronic files for archiving purpose needs to look at the entire lifecycle of the document or record. This not only includes the short term archival of documents at the preparer's facility, but also the cost associated with maintaining the documents at long-term storage facilities such as the National Archives or a data warehouse.

The factors that must be considered are as follows:

- Selecting documents for storage.
- Short-term document storage (at preparer's facility).
- Cost of transferring the document to a long-term storage facility, to include document conversion from one format to another, the media that will be used to transport the document, etc.
- Costs of maintaining the document in the long-term storage facility.

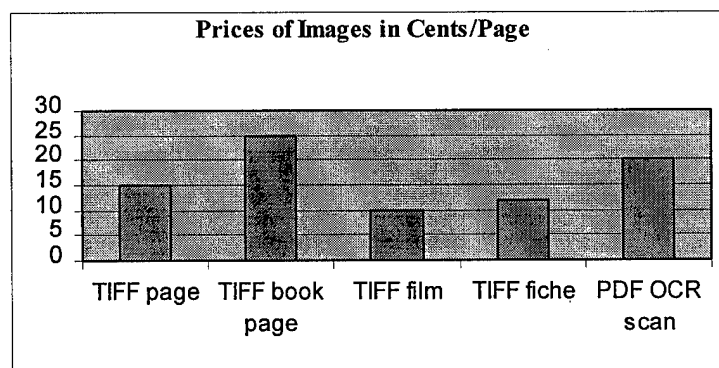
This short list assumes that the organization has already developed standards, methods and processes for creating electronic images, storing electronic files, and transferring electronic records between organizations. If these efforts have not already been accomplished by the organization responsible for the life cycle of the document, then the costs associated with these efforts must be factored in as well.

During the course of our research we collected data from organizations that market the service of digital image record creation. The following table provides cost information that was obtained from one organization that is currently involved with a large-scale document-imaging project that is creating electronic files for a government organization:

**TABLE 2. COST INFORMATION**

Method of Conversion	Format	Price
Single pages to electronic image	TIFF	10-16¢ /page
Bound book pages to electronic image	TIFF	22-25¢/page
Film image to electronic image	TIFF	10¢/page
Fiche image to electronic image	TIFF	12¢/page
Single page or from TIFF image	PDF with text at 96% OCR accuracy	20¢/page - \$2.10/page

Throwing out the high price (labor intensive – high percentage of text re-keyed textual information) the comparison is provided in the following graph:



## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

### FIGURE 2. IMAGE PRICES

These low-end prices are for material that does not require a lot of preparation and indexing and material that does not require special handling and can be feed automatically into the imaging equipment. Equipment costs, acquisition and maintenance, staff training, quality control, verification, metadata creation, and software costs are spread over a large business base.

Imaging from microfilm is almost entirely automated and has lower associated costs since the preparation has been previously done when material originally microfilmed. For projects that are intended for both archive and access, the industry recommends producing not only the scanned TIFF image, but in the same process, feed the image to a computer-output microfilm (COM) machine and produce microfilm for archival purposes. The rule of thumb is that if you want *access*, use the electronic image, but if you want *preservation*, then use microform.

A second cost sample was obtained from the final report published by Cornell University on the *Digital to Microfilm Conversion: A Demonstration Project*. Cost figures were provided in this report not only for the conversion of digital images to microfilm, but also for the conversion of microfilm to digital images. Yale University (Project Open Book) conducted the microfilm-to-digital project at the same time as the Cornell University project, and information was shared between the two institutions. The following cost information was provided in the *Cornell University Digital to Microfilm Conversion: A Demonstration Project 1994-1996, Final Report* to the National Endowment for the Humanities by Anne R. Kenney:

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

**TABLE 3. PRODUCING DIGITAL IMAGES FROM PAPER VS. MICROFILM**

	Cornell: Time & Costs			Yale: Time & Costs		
Process	Mean Time	\$/Bk	\$/Image	Mean Time	\$/Bk	\$/Image
Preparation	78.8 min	\$20.20	\$0.094	5.3 min	\$1.36	\$0.006
Scanning						
Auto	56.1 min	\$14.38	\$0.067	38.1 min	\$9.77	\$0.045
Manual	73.2 min	\$18.76	\$0.087			
Indexing	8.6 min	\$2.20	\$0.010	29.9 min	\$7.66	\$0.035
Other	5.2 min	\$1.33	\$0.006	19.2 min	\$4.92	\$0.023
Sub Total: Process						
Auto	148.7 min	\$38.11	\$0.18	92.5 min	\$23.71	\$0.110
Manual	165.8 min	\$42.49	\$0.020			
Equipment	Mode			Capacity		
	Auto	\$14.30	\$0.066	High	\$24.51	\$0.113
	Manual	\$17.40	\$0.080	Low	\$31.32	\$0.145
Total: Process/Equip	Mode			Capacity		
	Auto	\$52.41	\$0.24	High	\$48.22	\$0.22
	Manual	\$59.89	\$0.28	Low	\$55.03	\$0.26

The following table provides direct comparison of the two data samples:

**TABLE 4. DIRECT COMPARISON**

	Sample 1	Sample 2
Digital Image per Page	\$0.13 <sup>1</sup>	\$0.25 <sup>2</sup>
Digital Image per Bound Book <sup>3</sup>	\$54.00	\$53.88 <sup>2</sup>

Notes:

1. This was the average cost reported in Sample 1, which ranged from \$0.10 - \$0.16/page.
2. These figures are an average of the Cornell University and Yale University costs, which ranged from \$0.22 to \$0.28.
3. Both figures are based on a 216-page book.

There are several factors that can explain the difference in the Digital Image per Page cost between these two sets of data. One is that the Cornell and Yale University studies were established as demonstration or proof-of-concept projects, while the Government project was a competitive bid effort. Another explanation could be that the University projects were accomplished in-house, while the Government project was contracted out. Regardless of the differences, these examples serve the purpose of providing cost information that can be used for planning purposes. For more detailed information, the Cornell University report can be downloaded from the WWW at **Error! Bookmark not defined..**

Within in the DoD the vendor or outsourcing agent is the Defense Automated Printing Service (DAPS). DAPS is responsible for document automation and printing within the Department of Defense, encompassing electronic conversion, retrieval, output and distribution of digital and hardcopy information.

DAPS sees conversion as one of the most important services to their customers over the next few years. The DoD has issued strategic goals and objectives that require the DoD to transition into paperless environments. The DoD is looking to DAPS for support in moving from a paper-based to a digital environment, including raster scanning, engineering drawings, Tag Image File Format (TIFF), Group 4 format, quality Assurance, CD-ROM and WORM, SGML to HTML, hyperlinked PDF and much more.

## 5.6 Migration Strategies

As the operating environments of digital archives change, it becomes necessary to migrate their contents. There are a variety of migration strategies for transferring digital information from systems as they become obsolete to current hardware and software systems so that the information remains accessible and usable. No single strategy applies to all formats of digital information and none of the current preservation methods is entirely satisfactory. Migration strategies and their associated costs vary in different application environments, for different formats of digital materials, and for preserving different degrees of computation, display, and retrieval capabilities. The general rule of thumb appears to be plan for your migration efforts to cost between 50 – 100% of the cost to create the original digital image document.

Methods for migrating digital information in relatively simple files of data are quite well established, but the preservation community is only beginning to address migration of more complex digital objects. Additional research on migration is needed to test the technical feasibility of various approaches to migration, determine the costs associated with these approaches, and establish benchmarks and best practices. Although migration should become more effective as the digital preservation community gains practical experience and learns how to select appropriate and effective methods, migration remains largely experimental and provides fertile ground for research and development efforts.

One migration strategy is to transfer digital materials from less stable to more stable media. The most prevalent version of this strategy involves printing digital information on paper or recording it on microfilm.

Retaining the information in digital form by copying it onto new digital storage media may be appropriate when the information exists in a "software-independent" format as ASCII text files or as flat files with simple, uniform structures.

Copying from one medium to another has the distinct advantage of being universally available and easy to implement. It is a cost-effective strategy for preserving digital information in those cases where retaining the content is paramount, but display, indexing, and computational characteristics are not critical. As long as the preservation community lacks more robust and cost-effective migration strategies, printing to paper or film and preserving flat files will remain the preferred method of storage for many institutions and for certain formats of digital information.

Another migration strategy for digital archives with large, complex, and diverse collections of digital materials is to migrate digital objects from the great multiplicity of formats used to create digital materials to a smaller, more manageable number of standard formats that can still encode the complexity of structure and form of the original. A digital archive might accept textual documents in several commonly available commercial word processing

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

formats or require that documents conform to standards like SGML (ISO 8879). Databases might be stored in one of several common relational database management systems, while images would conform to a tagged image file format (TIFF) and standard compression algorithms (e.g., JPEG).

Changing format as a migration strategy has the advantage of preserving more of the display, dissemination, and computational characteristics of the original object, while reducing the large variety of customized transformations that would otherwise be necessary to migrate material to future generations of technology. This strategy rests on the assumption that software products, which are either compliant with widely adopted standards or are widely dispersed in the marketplace, are less volatile than the software market as a whole. Also, most common commercial products provide utilities for upward migration and for swapping documents, databases, and more complex objects between software systems. Nevertheless, software and standards continue to evolve so this strategy simplifies but does not eliminate the need for periodic migration or the need for analysis of the potential effects of such migration on the integrity of the digital object.

Use of one of the evolving interchange standards, such as the Basic Image Interchange Format (BIIF) or Electronic Document Interchange Standard (EDIS) allows for the receipt of images in many different formats which are converted into one robust format. Having only one format that will handle all types of images simplifies the migration issue to handling of only one format.

BIIF is based on the National Imagery Transmission Format Standard (NITFS) developed by the DoD and adopted by North Atlantic Treaty Organization (NATO). The BIIF is the basis for a new standards activity within ISO/IEC JTC1/SC24 to add a new part 5 to the International Standard for Image Processing and Interchange (IPI) (ISO 12087-5, 1998)

BIIF specification provides such a common basis for storage and interchange of images and associated data among existing and future applications. BIIF supports interoperability by providing a data format for shared imagery and an interchange format for images and associated imagery data. The documentation provides a detailed description of the overall structure of the format, as well as specification of the valid data content and format for all fields defined within a BIIF file. BIIF provides a data format container for raster, symbol, and text data, along with a mechanism for including image-related support data.

BIIF satisfies the following requirements:

- Allow diverse applications to share imagery and associated data.
- Allows an application to exchange comprehensive information to users with diverse needs or capabilities, allowing each user to select only those data items that correspond to their needs and capabilities.
- Minimizes preprocessing and post processing of data.
- Minimizes formatting overhead, particularly for those applications exchanging only a small amount of data and for bandwidth-limited systems.
- Provides a mechanism to interchange Programmer's Imaging Kernel System (PIKS) (Part 2 of ISO 12087) image and image-related objects
- Provides extensibility to accommodate future data, including objects. As BIIF becomes more capable through extension and the addition of new data, objects and data relationships, concepts and features of 12087-3 (Image Interchange Format [IIF]) may be considered as a more appropriate method of growth. This is to facilitate a growth path from BIIF to IIF.

In BIIF, data interchange between disparate systems is potentially enabled by a translation process. Using BIIF, each system must be compliant with only one external format that will be used for communication with all other participating systems. When BIIF is not used as a system's native internal format, each system will translate between the system's internal representation for imagery and the BIIF format. A system from which data is to be transferred has a translation module that accepts information structured according to the system's internal representation for images and related imagery data, and assembles this information into the BIIF format. The receiving system will reformat the BIIF data, converting it into one or more files structured as required by the

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

internal representation of the receiving system. Each receiving system can translate selectively and permanently store only those portions of data in the received BIIF that are of interest. A system may transmit all of its data, even though some of the receiving systems may be unable to process certain elements of the data.

Profiles of BIIF will be established as International Standardized Profiles (ISP) through the ISO process (ISO/IEC TR 10000).

EDIS is a voluntary standard for electronic document interchange among Executive Branch agencies, which review electronic images of documents. The standard governs both document metadata and document images that are to be exchanged for purposes of coordinating review, as well as minimum transfer metadata. This Standard is designed solely to provide specifications for the interchange of electronic documents and related information between systems. The Standard was developed by the Declassification Program Managers Council (DPMC) Automation Working Group (AWG) and The George Washington University Declassification Productivity Research Center (DPRC) for the declassification community.

## 6 Conclusions

1. Access and response to Freedom of Information Act (FOIA) requests are facilitated through electronic digitalization of records.
2. No de jure standard for digital images has reached the desired maturity level for archival purposes.
3. The hardware and software technology required for the use of digital images changes rapidly.
4. Migration costs associated with archiving of digital images of textual material are unknown.
5. The anticipated high cost associated with long-term maintenance of digital image records mandates careful screening and selection of only the most valuable digital imaged records to be accessioned into the National Archives.
6. Metadata standards have been developed, but no one standard has emerged as the most universally accepted standard for electronic image records.
7. Tag Image File Format (TIFF) and Portable Document Format (PDF), both de facto standards, are the most widely used formats for text records.
8. Joint Photographic Experts Group (JPEG), a de jure standard, is the most widely used compression standard.
9. The use of proprietary standards for producing and storing images is much more common than the use of official standards.
10. Organizations will continue to use the proprietary imaging formats due to the costs involved.
11. The key roadblock to a successful digital imaging program is the high costs associated with the program and the lack of management understanding to the need for appropriate funding in the area.
12. The lack of a format standard is no longer seen as a major issue
13. A united government voice was needed, with strong NARA leadership and a means of sharing data.

The following phased implementation approach received general acceptance at the DoD-NARA Scanned Images Standards Conference:

1. Manage the process (records management, management and policy)
2. Study, plan, gather information through cost/benefit analysis of entire life-cycle (especially document preparation, searching, and migration).
3. Pick an interim standard during step 2, which will be accepted and supported by DoD and NARA – this will enable the cost-benefit analysis to be conducted.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

4. Practice migration and preservation while documents are in active use

## 7 Recommendations

1. Image electronic digital material in the most stable formats available preferably using the latest version, but no more than two generations prior to the latest. (e.g., for TIFF image produced in January 1999 that would be TIFF version 6, 5 or 4)
  - a. Image personnel records using TIFF for archiving, TIFF or PDF formats for access. Convert all current imaged records to one standardized TIFF format.
  - b. Image declassified records using TIFF for archiving, TIFF 6 or PDF formats for access. Convert declassified versions of historically significant records to paper, microfilm or ASCII formats.
  - c. Image manuals, standards, directive type material using TIFF, ASCII and ASCII SGML or XML tagged files for archiving. Use PDF, HTML or XML formats for dissemination.
2. Plan and budget for migration of digital images every 3-5 years with of cost equivalent to 50 – 100% of the costs associated with original imaging project.
3. Convert documents that require long-term preservation from application format to an image format.
4. Develop standard set of access metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set.
5. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.
6. Work with NARA to:
  - a. Establish criteria for selection of digital images for accessioning in the National Archives
  - b. Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.
  - c. Establish guidelines describing metadata that must accompany digital image when submitted for archival accessioning
  - d. Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
  - e. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
  - f. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.

## 8 References

The Applied Technologies Group, Enhancing Applications with Imaging Technology, The Technology Guide Series, 1998.

"Automatic Assessment of Document Image Quality," The Communicator, Vol. III, No. 1, Summer 1998, [http://dprc.seas.gwu.edu/dprc5/current\\_news/communicator/communicator\\_sum98.htm](http://dprc.seas.gwu.edu/dprc5/current_news/communicator/communicator_sum98.htm).

Behavioral Computational Neuropsychology (BCN) Group, Founding Committee, First General Announcement Proposing a Knowledge Processing Manhattan Project, May 24, 1998.

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Carlin, John W., Statement on the Report of the Electronic Records Work Group Of the National Archives and Records Administration (NARA), 21 September 1998. **Error! Bookmark not defined..**

Carlin, John W., "The Uncertain Future of the Past," Washington Post, August 23, 1998, p. C3.

Carson, Steve, "Basic Image Interchange Format (BIIF)," GSC Associated Inc., **Error! Bookmark not defined..**

Cohen, Edmund, "Declassification and Records Management;" U.S. Central Intelligence Agency; Presentation, August 31, 1995.

Declassification Productivity Research Center, The George Washington University, Declassification Products, [http://dprc.seas.gwu.edu/dprc5/declassification\\_products/products.htm](http://dprc.seas.gwu.edu/dprc5/declassification_products/products.htm).

Declassification Productivity Research Center, The George Washington University, Declassification Technologies, [http://dprc.seas.gwu.edu/dprc5/declassification\\_tech](http://dprc.seas.gwu.edu/dprc5/declassification_tech).

Declassification Productivity Research Center, The George Washington University, Integrated Declassification for Executive Agencies (IDEA) System, [http://dprc.seas.gwu.edu/dprc5/declassification\\_products/idea.htm](http://dprc.seas.gwu.edu/dprc5/declassification_products/idea.htm).

DoD 5015.2-STD "Design Criteria Standard for Electronic Records Managements Software Applications," <http://jtc.fhu.disa.mil/recmgt/rma-ps/index.html>

EDIS, [http://dprc.seas.gwu.edu/dprc5/declassification\\_products/EDIS\\_ver1.2.htm](http://dprc.seas.gwu.edu/dprc5/declassification_products/EDIS_ver1.2.htm).

Electronic Imaging Standards for Archiving Records, report prepared by Logicon Communications Group for Deputy Assistant Secretary of Defense, Information Management, May 31, 1997.

Gray, Douglas E., "Preparing Graphics for The Web – Which Format to Use," May 25, 1997 **Error! Bookmark not defined.**

Guidelines on the Management of Electronic Records from Office Systems. "Chapter 6: Transfer to the PRO," pg. 43-51, **Error! Bookmark not defined..**

Hickok, Gene J., "Media Life-Cycle Costs and Integrity;" National Media Laboratory; Presentation.

Information technology – Computer Graphics and image processing – Image Processing and Interchange (IPI) - Functional Specifition – Part 5: Basic Image Interchange Format (BIFF), ISMC – ISO?IEC 12087-5:1998(E), 1 DECEMBER 1998. <http://164.214.2.51/ntb/baseline/docs/biif/index.html>

JPEG: **Error! Bookmark not defined..**

Kenney, Anne R. and Rieger, Oya Y., "Using Kodak Photo CD Technology for Preservation and Access," A Guide for Librarians, Archivists, and Curators; Department of Preservation and Conservation, Cornell University Library, May 1998.

National Library of Australia, "Preserving Access to Digital Information," **Error! Bookmark not defined.**

"Preserving Digital Information, Final Report and Recommendations" by the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group, May 20, 1996, **Error! Bookmark not defined..**

Prueitt, Paul S., "Induction, Deduction and a New Form of Machine Intelligence;" Declassification Productivity Research Center, The George Washington University; April 8, 1998.

Prueitt, Paul S., "Measurement, Categorization and Bi-level Computational Memory;" Declassification Productivity Research Center, The George Washington University; June 1, 1998.

Puglia, Steven and Roginski, Barry; "NARA Guidelines for Digitizing Archival Materials for Electronic Access;" National Archives and Records Administration; January 1998.

Raster Graphic Interchange Standards **Error! Bookmark not defined..**

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Research Library Group, "The RLG Worksheet for Estimating Digital Reformatting Costs," Error! Bookmark not defined.

"A Strategic Policy Framework for Creating and Preserving Digital Collections," Error! Bookmark not defined.

"Technical Recommendations for Digital Imaging Projects," Error! Bookmark not defined.

The VRML Repository, Error! Bookmark not defined.

"Why all the argument about file formats?" Error! Bookmark not defined..

UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

## APPENDIX A - FORMATS

Information In Table Based On Data From OII Raster Formats

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
BIIF	Complex raster imagery interchange standard	ISO/IEC 12087-5	BIIF is based on NITFS developed by the DoD and adopted by NATO. It is part 5 to the International Standard for Image Processing and Interchange (IPI). Has sections for raster images and vector graphics/text composite images	Accepted standard for imagery. Used by NIMA and other US and NATO intelligence organizations.
BMP	Proprietary raster image file interchange format	Microsoft's Window Device Independent Bitmap format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
CUT	Proprietary raster image file interchange format	Media Cybernetic's Dr. Halo graphic format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
DDES	Raster image interchange on magnetic tape	ISO 10755, 10756, DIS 10757, 10758, 10759	Based on ISO 1001 (ANSI X3.27) - Volume and File structure on magnetic tape.	Widely implemented in graphic arts industry - but outside of "big four" graphics firms is virtually unused. Dissatisfaction due to time it takes to translate between vendors' formats. PDF and TIFF have undermined need for DDES.
EDIS	Raster image interchange standard	DPMC AWG and DORC	Uses TIFF 6 format. Voluntary standard for electronic document interchange among Executive Branch agencies to review electronic images of documents.	US agencies involved with Declassification Project
Fax Groups 3 & 4 (CCITT)	Encoding for facsimile transmission (CCITT)	CCITT (now ITU) T.4 and T.6	Lossless compression based on run-length and Huffman encoding.	Accepted standard for facsimile. Commonly used for transmitting and archiving images that are included in printed or hypertext documents. Compression of images for transfer, used with TIFF-F.

UNCLASSIFIED

GA22F042

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
Fractal transform coding	Raster image compression method	Patented - Iterated Systems, Inc.	Image compression. Intraframe compression in excess of 10,000:1 and interframe of 1,000,000:1.	Slow, computationally intensive. Widely used in Iterated Systems produces and Microsoft Encarta CD-ROM multimedia product.
GEF	Raster image file format	Telegrafix	Alternative to GIF	Alternative to GIF that does not use LZW.
GIF	Raster image file format	CompuServe Inc. (royalty-free limited-use license to users)	Graphic interchange, irrespective of system. Stored as pixels with RGB color values. GIF files start with signature (version number, etc.), screen definition, and global color scale of GIF-generator hardware. Compressed according to the LZW algorithm.	Used for e-mail of images for viewing on screen. Widely supported by many applications.
ILBM	Proprietary raster image file interchange format	Commodore Amiga Interleaved Bitmap format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
IMG	Proprietary raster image file interchange format	Digital Research's GEM Image format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
IPI-IIF	Data format and gateway functionality for raster image interchange	ISO/IEC JTC1/SC24  ISO/IEC 12087 -1 common architecture -2 PIKS -3 IIF -5 BIIF  ISO/IEC 12089	IPI data format goes beyond the capabilities of TIFF by allowing more than two dimensions (e.g. XYZ and T) for the "sampling space."	Standard still under development. Generic image standard. Commercial implementation became avail in 1996. National Institute of Standards (NIS) developed PIKS test suite with Sun Microsystems. A consortium of medical system vendors is implementing IIF and CEN TC251 W64 is evaluating its use as part of MEDICOM. DoD and NATO plan to use the BIIF format. Allows up to 5 dimensions - allowing time varying images and multi-channel images useful for geosciences, fluid analysis, medical imaging, etc. Not likely to displace TIFF for two-dimensional graphic images.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
JBIG	Bilevel raster image encoding standard	ITU-T (CCITT) and ISO/IEC JTC1/SC29  ITU-T Rec. T.82/ISO/IEC 11544	Defines a bit-preserving (lossless) compression method for coding. Developed by the Joint Bi-level Images experts Group (JBIG). Compression 1.1 - 1.5 times as great as CCITT Groups 3 and 4. JPEG more commonly used for greyscale and color images.	Has not been as well adopted.
JPEG (SPIFF)	Continuous-tone (photographic) still image compression standard	ITU-T (CCITT) ISO/IEC JTC1/SC29  ISO/IEC 10918  ISO/IEC DIS 14495-1	For compression of monochrome or full-color, greyscale, digital still images. Options allow use of lossy (DCT) or lossless (predictive) algorithms. JPEG defines bitstream of encoded compressed image data (communication line). For data processing the bitstream needs to be encapsulated in a file format like SPIFF.	Many implementations both HW and SW including public domain SW.
LZW	Compression algorithm	IBM and UNISYS Developers pay royalty.	Used with GIF	Used with GIF.
MSP	Proprietary raster image file interchange format	Microsoft Window Paint format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
ODA RGCA	Raster graphics within ODA documents	ISO/IEC JTC1/SC18, ITU (CCITT) and ECMA  ITU-T Recommendation T.417  ISO 8613-7	Office Document Architecture (ODA) allows raster graphics encoded using CCITT Groups 3 and 4 fax, JBIG and JPEG specification to be embedded with ODA documents, and controls the scaling and positioning of such diagrams within formatted text.	Use restricted by limited acceptance of ODA.
PBM	Proprietary raster image file interchange format	Poskanzer's Portable Bit Map utilities	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
PCX	Proprietary raster image file interchange format	Zsoft's PC Paintbrush format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
Photo CD	Raster images stored on CD-ROM	Proprietary standard patented by Eastman Kodak Company.	Means of storing high quality digital color images, captured from continuous-tone film, as a digital signal on a CD-ROM disc.	Professional photography market.
PICT	Proprietary raster image file interchange format	Apple's Picture format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
PNG	Compression Algorithm	IETF	24-bit lossless algorithm.	Royalty-free replacement of LZW.
PNG	Raster image file format	W3C PNG Specification Version 1.0	Extensible file format for lossless portable, well-compressed storage of raster images. Patent-free replacement for GIF and can also replace many common uses of TIFF.	Expected to replace GIF as the main raster file format alongside the JPEG standard developed by ISO.
PNTG	Proprietary raster image file interchange format	Apple's MacPaint format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
SCR	Proprietary raster image file interchange format	Microsoft's Screen Capture format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
SUN	Proprietary raster image file interchange format	Sun Raster Files format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
TGA	Proprietary raster image file interchange format	Truevision's TARGA format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.

UNCLASSIFIED

1 June 1999

## Imaging Standard Support Task

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
TIFF	Raster image interchange format	Proprietary format developed by Aldus Corporation (now owned by Adobe) and Microsoft  Per Basil Manns (LC) MS 53 is standard for header format.	Tag-based file descriptor that can characterize almost any form of 2D raster data using either ASCII or binary coding. Private tags may be used. Standard TIFF allows the use of PacBits, LZW, Group 3 or 4 Fax and JPEG compression schemes within transmitted images.	Widely implemented in desktop publishing (DTP) and desktop applications. Used by scanners and digital image interchange format and by some applications as memory format. Wide variation between requirement for different applications has led to partial implementations, resulting in many valid TIFF files which cannot be read. The situation is being addressed by the ANSI IT8.8/TC130 activity, which aims to standardize the TIFF subset required for Graphic Technology applications.
TIFF/IT	Standardized raster image file interchange format	ANSI IT8  ISO/TC130/WG2  ISO 12639	Second generation standard which will provide functionality of ANSI IT 8.1, 8.2 and 8.5 to media other than magnetic tape; provide a data format for high resolution edge information; provide a format standard for the subset of TIFF which is appropriated for prepress applications. Designed so that any fields additional to TIFF 6.0 take default values equivalent to TIFF 6.0 practice, so that existing implementations should already be compatible with it.	Too soon to tell (1998 standard). Intended that the ITFF/IT will reduce the variation between TIFF implementations which has led to current unreadable issues with TIFF files.
WMF	Proprietary raster image file interchange format	Microsoft's Windows Metafile format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
WPG	Proprietary raster image file interchange format	WordPerfect's graphic format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.
XBM	Proprietary raster image file interchange format	X-Windows Bit Map	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.

UNCLASSIFIED

UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
XWD	Proprietary raster image file interchange format	X-Windows Dump screen window image format	Developed by SW supplier to move raster graphics between software packages.	Widely implemented in DTP and desktop application of the company.

UNCLASSIFIED

GA22F042

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

## Vector Graphic Interchange Standards

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
CDR	Proprietary vector image file interchange format	Corel Draw format	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.
CFF2	Vector and text interchange by modem or diskette	None	Enables senders/receivers of data needed for the manufacture of dieboards or artwork to overcome language barriers. Only limited set of ASCII characters allowed. KERMIT used as communication protocol.	Must be obtained from one of the main suppliers of CAD-CAM systems for packaging industry.
CGM	Metafile for the storage and transfer of picture description information	ISO/IEC JTC1/SC24 ISO/IEC 8632	Machine and operating system independent interchange format. Provides elements to present geometric graphics and raster graphics. Consists of a functional specification and multiple encodings for different purposes. The only standard for graphical database specification designed to serve a wide range of applications. CGM is a static picture-capture metafile.	Basis of Geometric Graphics Content Architecture part of ODA. Significant market. Graphic exchange format with the US DoD CALS initiative. Many graphics packages today capable of generating and/or interpreting CGM files. Since 1996 tools for transmitting CGM illustrations as part of an Internet file set have been developed. CGM is now a recognized MIME data type and moves are being made to introduce it as one of the default set of image formats for use within HTML documents.
CGRM	Terminology and general model for describing the inter-relationships between graphics standards	ISO/IEC JTC1/SC24 ISO/IEC 11072	Based on a layered model, and thus has a superficial similarity to the ISO OSI 7-layer model. CGRM has only five layers.	Intended for developers of computer graphics standards.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
DDES2-IADD	Interchange of die-cutting specifications by modem or diskette	ANSI IT8.6	Communication protocol for CAD-CAM systems for the design of folding boxes of both solid and corrugated packages on one side and automated CNC-controlled dieboard cutting machines (laser and jigsaws) on the other.	Protocol is mainly used in the US and Canada.
DRW	Proprietary vector image file interchange format	Micrografix Designer format	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.
DXF	Vector graphics principally used for CAD drawings	Autodesk Inc.	Allows the transfer of AutoCad 3-D drawings between CAD and illustration applications.	Used for interchange between CAD and other vector drawing packages, particularly on PC and UNIX computers. Most PC drawing and illustration software supports the import and export of this format.
GEM	Proprietary vector image file interchange format	Digital Research's GEM metafile format	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.
GKS	Vector graphic interchange	ISO/IEC JTC1/SC24 ISO/IEC 7942	Machine, language, operating system, and device-independent specification of a set of services for displaying and interacting with 2D pictures. Subroutine library rather than a data format.	Often supplied as an integral part of the operating system of a graphics workstation. Laid the groundwork for a common terminology and common concepts across the whole of computer graphics, and also has provided guidelines for the development of on-chip functions.
GKS-3D	Three dimensional vector graphics	ISO/IEC JTC1/SC24 ISO 8805	Machine, language, operating system, and device-independent specification of a set of services for displaying and interacting with 2D and 3D pictures.	Use of GKS-3D not as wide as GKS; for most 3D applications structuring is an important and required functionality which is only supported to a nesting level of one within GKS-3D.
HPGL	Proprietary vector image file interchange format	Hewlett-Packard Graphic Language	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
IGES	Geometric (vector) and non-geometric entities	NBS then ANSI	IGES is the principal standard for the exchange of product definition data in the US. There are two different data formats - ASCII and binary.	Most CAD-CAM packages provide pre- and post processors for IGES. As a US DoD initiative was compulsory for suppliers to the US defense industry. Profiles for mechanical engineering drawings, pipework layouts, circuit board design.
ISO 10303-STEP	An international standard for the computer-interpretable representation of product data and its exchange.	ISO TC184/SC24 ISO/IEC 10303-42	Deals with the representation of information found in vector diagrams.	Used with the STEP community as the basic mechanism for interchanging vector information.
ODA GGCA	Vector graphics with ODA documents	ISO/IEC JTC1, ITU (CCITT) and ECMA ITU-T Recommendation T.418 ISO 8613-8	Architecture allows geometric graphics encoded using the binary encoding method provided by the 1987 version of CGM to be embedded with ODA documents, and controls the scaling and positioning of such diagrams within formatted text.	See CGM
PHIGS	Storage and interchange of 3D geometric models	ISO/IEC JTC1/SC24 ISO/IEC 9592	Machine, language, operating system, and device-independent specification of a set of services for displaying and interacting with 3D geometric model.	Delivered as an integral part of many high performance graphical workstations.
PIC	Proprietary vector image file interchange format	Lotus 1-2-3 graphic interchange file	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.
PICT	Proprietary vector image file interchange format	Apple's Picture format	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.
PLOT	Proprietary vector image file interchange format	Unix Plot format	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.

# UNCLASSIFIED

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Name	Area covered	Sponsoring body and standards	Characteristics/ descriptions	Usage (Market segment and penetration)
SET	Storage and interchange of geometric data	AFNOR	Standard for the interchange of engineering data with France and within European multinational projects. Has been superseded by STEP.	Standard for several of the EEC co-operative projects such as Airbus and the Hermes Space Shuttle. Heavily used with the European aircraft industries, but not much outside this area.
VDAFS	Interchange of surface related engineering data	DIN	Standard for the interchange of engineering data within Germany.	Mostly used in the German automobile industry.
WMF	Proprietary vector image file interchange format	Microsoft's Windows Metafile format	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.
WPG	Proprietary vector image file interchange format	Wordperfect's Graphic file format	Developed by SW supplier to move vector graphics between software packages.	Widely implemented in DTP and desktop applications of the company.

## APPENDIX B - SURVEY

In October 1998, a survey focusing on the current usage of electronic images and archives was sent to thirty-five Government agencies. A 71% return ratio was obtained from 25 respondents, suggesting the high level of interest in the area of using, establishing and maintaining electronic archives.

A copy of the survey is provided at the end of this Appendix.

### Survey Analysis:

There were several key issues that were identified from this survey that either confirmed the current thought, or provided answers to some key questions. These significant findings were as follows:

- 1) *A vast majority of respondents (92%) believed that either NARA or OSD should provide standards on how records should be stored in an electronic archive.*
- 2) *A vast majority of the respondents (88%) would like to receive direction on how to establish, implement and maintain an electronic archive.*
- 3) *A smaller percentage of respondents (72%) felt that either NARA or OSD should mandate these standards.*

These three related findings illustrate the importance that the DoD agencies place on the proper handling of electronic records, and point out the necessity for guidance in the areas of establishing electronic archives.

The OSD and NARA have made considerable efforts in establishing standards and providing recommendations on the handling of electronic records. However, it has become almost impossible to keep up with the changing technology arena. New hardware and software products are hitting the market every day, making both the existing de facto and de jure standards obsolete almost as soon as they are released.

This report is a prime example of the efforts the OSD and NARA are currently undertaking in their attempt to address and resolve these issues. Until guidelines can be established, the Information Officers responsible for managing the DoD's electronic records should keep themselves abreast of the current state of technology, and with the current efforts that are underway in the area of establishing and maintaining electronic archives. There are a number of different studies currently underway that provides a constant stream of information on this topic. The list of WWW References found in Appendix G of this document can be used as reference material for establishing procedures on properly setting up and maintaining an electronic archive that will ensure the digital information produced today will remain accessible in the future.

- 4) *The use of Optical Disks by Government organizations is high. More respondents stated that they were using Optical Disks to store electronic records than any other type of media.*

This issue is of great concern to the National Archives and Records Administration. The reason for this is that the optical disks are notorious for having incompatibility problems. Unlike CD-ROM disks, the format for the optical disks was never standardized. This resulted in each manufacture, and in some cases, each model from a particular manufacture having a different format and therefore being incompatible. The concern is that when these optical disks are sent to the National Archives and Records Administration for permanent storage, the hardware will not be available to read the disks.

An excellent example of this is a statement provided from a contract manager for a large imaging project:

"When I buy a juke box, I make sure that I buy enough optical disks to completely populate the juke box even if they are not required to support the task. The reason for this is that chances are high that if I go back to the same vendor to buy additional optical drives in six months, the disks will no longer be available."

Current requirements are that electronic records transferred to the NARA will be on either 7 or 9 track open-reel magnetic tape, or on 18 track 3480-class tape cartridge.

The NARA does not expect that this requirement can be enforced 100% of the time, and anticipates that records will be provided on any type of media conceivable. Past experience has shown that when the records are scheduled for transfer to the archives, the records will be delivered in its current state, and federal law requires NARA to accept all of these records.

The NARA is attempting to establish other types of media that will be accepted. However, due to the incompatibility problems, optical disks are one type of media that will probably never be added to the list of recommended media types.

Government organizations that are currently using optical disks for storage of electronic files should keep this in mind. If there is ever a requirement to migrate the information currently on optical disks to another media, another optical disk solution may not be the most optimum long-term solution. The choice of CD-ROM or other types of media may be better suited for transferring information to the NARA, and could prevent the organization from having to take another migration step sometime in the future.

**5) *Quite frequently, electronic files are being stored in the electronic archive in the format in which they were created. 8 out of 11 responding organizations stated that they used this method at least part of the time.***

This is another issue that is of great concern to both the NARA and the OSD. Electronic records that are stored in Microsoft Word format or Word Perfect format may not even be readable in the near future. Also, there is an issue with migrating these proprietary formats from one version of the product to another. This migration can become a lengthy, time consuming process, and there is no guarantee that the reformatted product will be identical to the original.

If electronic records are going to be maintained, then the optimal solution is to store the electronic file as an electronic image. This image should be an exact replication of the original document, and is the most appropriate method for the long-term preservation of a record.

**6) *A majority (60%) of the responding organizations currently plans on implementing either a Document Management System and/or a Records Management System in the near future.***

In November 1997, the Assistant Secretary of Defense for Command, Control, Communications and Intelligence issued the Design Criteria Standard for Electronic Records Management Software Applications (DOD 5015.2-STD). This standard sets forth mandatory baseline functional requirements for Records Management Application (RMA) software used by DoD Components in the implementation of their records management programs. In November 1998, the NARA endorsed the use of this standard for Federal agencies.

This standard should be used by any Federal agency that is planning on implementing a records management system. As of 8 March 1999, the following Records Management systems have been approved by the DISA, Joint Interoperability Test Command as being compliant with the DoD 5015.2 standard:

- ForeMost 6.3 by Provenance Systems Inc.
- ForeMost 7.0 by Provenance Systems Inc.
- TRIM Version 4.2 by Tower Software Corporation
- CS-CIMS Version 2.5.0.37 by DynSolutions Inc. (with ForeMost 6.3)
- Panagon Integrated Document Management (IDM) Version 4.2 by FileNET (with ForeMost 6.3)
- DOCS Open Version 3.7.2 by PCDOCS Inc. (with ForeMost 6.3)
- e.POWER Version 1.5 by Universal Systems Inc.
- RIMS Studio Version 7.1 by PSSoftware Solutions Limited
- DMX Version 1.1 by Eastman Software (with ForeMost 7.0 and Microsoft Exchange Server 5.0 Version 7.3.2.2.0)
- RecordsManager, Version 1.1 by IBM

**7) *Scanning is the most common method of storing a document in an electronic archive, and TIFF is the most common imaging format that is used to archive documents.***

This is not surprising, since these two items go hand in hand. Scanning is the one of the most cost efficient methods of producing a digital image, and most scanning software today are capable of producing a TIFF image.

However, there are several issues associated with the use of scanners and the use of TIFF. The first is that not all scanners are equal; some are capable of producing a sharper, clearer image than others. Also, not all images will be scanned at the same resolution, since this is a user-controlled function. One user may scan their documents at 600 dpi resolution, while another will use 300 dpi resolution.

In January 1998, the NARA issued the NARA Guidelines for Digitizing Archival Materials for Electronic Access. However, this document was prefaced with the statement that *"The Guidelines do not constitute, in any way, guidance to Federal agencies on records creation or transfer to the National Archives of the United States."*

It is recommended that the NARA and OSD develop a set of technical recommendations that can be used by DoD agencies in creating digital archives. This recommendation is further supported by the survey findings, where 92% of the respondents believed that either NARA or OSD should provide standards on how records should be stored in an electronic archive.

(This space intentionally left blank.)

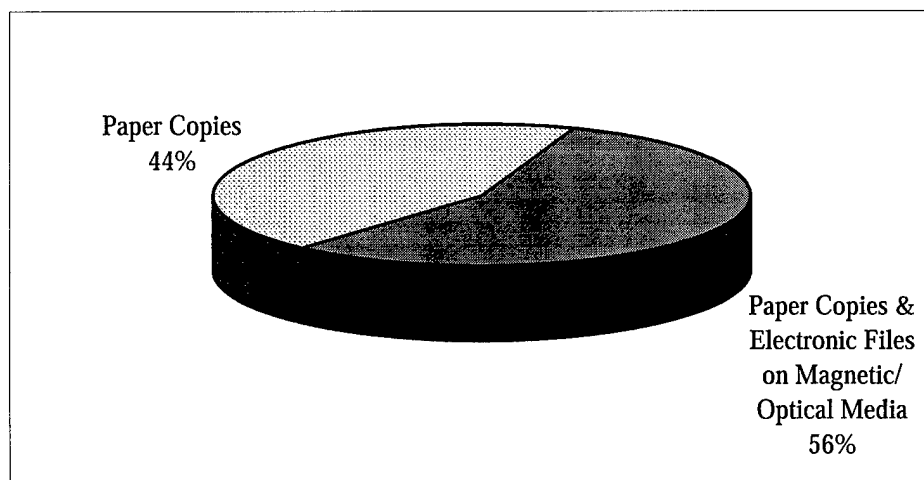
**Survey Results:**

*Question 1: Does your agency currently store or archive documents for future use?*

- a) 100% of the respondents currently store or archive documents for future use.

*Question 2: How are the documents archived?*

- a) 44% of the total respondents exclusively archive paper copies of documents stored in files or boxes.  
b) 56% of the total respondents currently store documents as either paper documents stored in files/boxes or electronic files stored on magnetic/optical media.  
c) 0% of the respondents archive documents exclusively using electronic media or methods.



**FIGURE B1 - HOW ARE DOCUMENTS ARCHIVED?**

*Question 3: If your organization maintains a document archive, what is its purpose?*

- a) 96% of the respondents answering this question stated that they maintain a document archive to retain records of documents for legal purposes.
- b) 88% of the respondents answering this question stated that the purpose of the document archive is to retain records of documents for historical purposes.
- c) 88% of the respondents answering this question stated that the purpose of the document archive is to allow information to be retrieved and shared throughout the organization.
- d) 83% of the respondents answering this question stated that federal, state or local laws require the document archive.

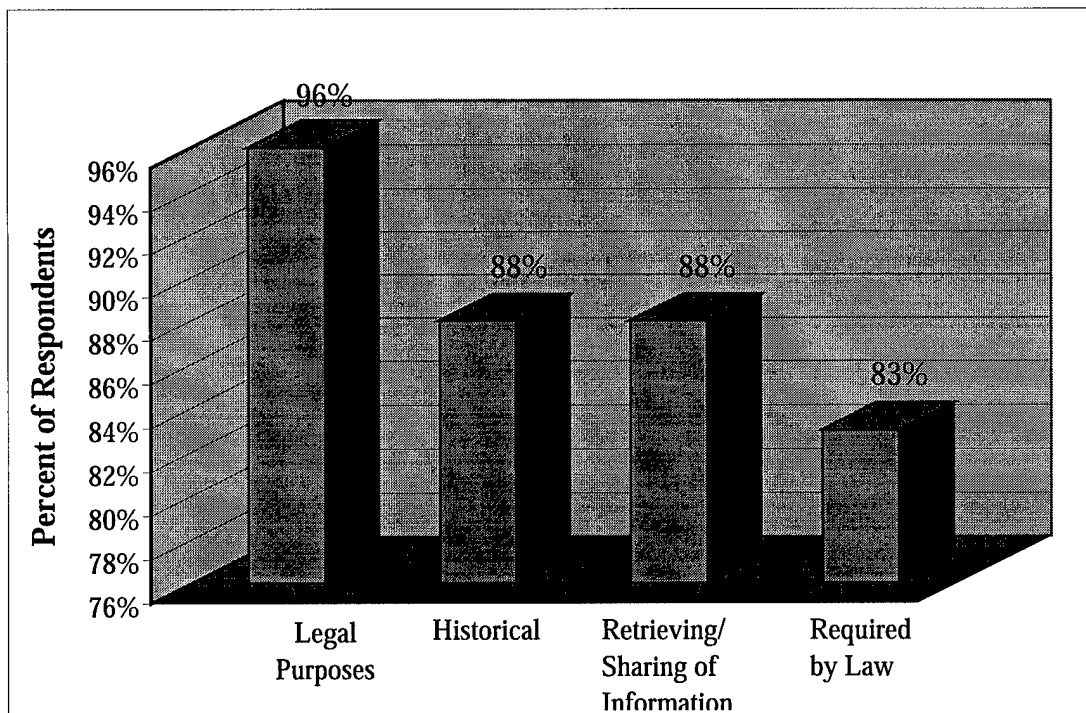
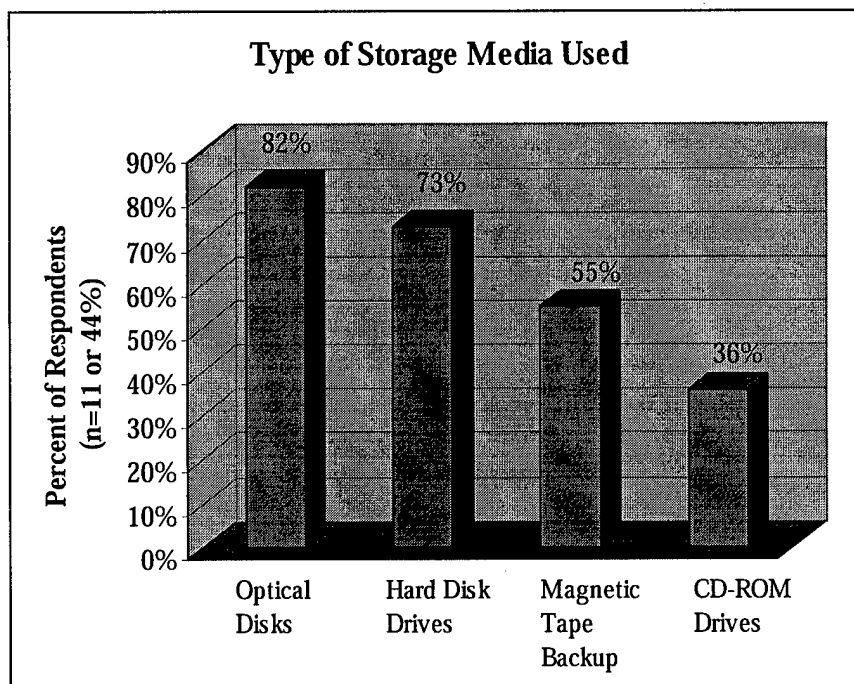


FIGURE B2 - PURPOSE OF DOCUMENT ARCHIVE

*Question 4: If your organization maintains an electronic archive, what type of storage media is used?*

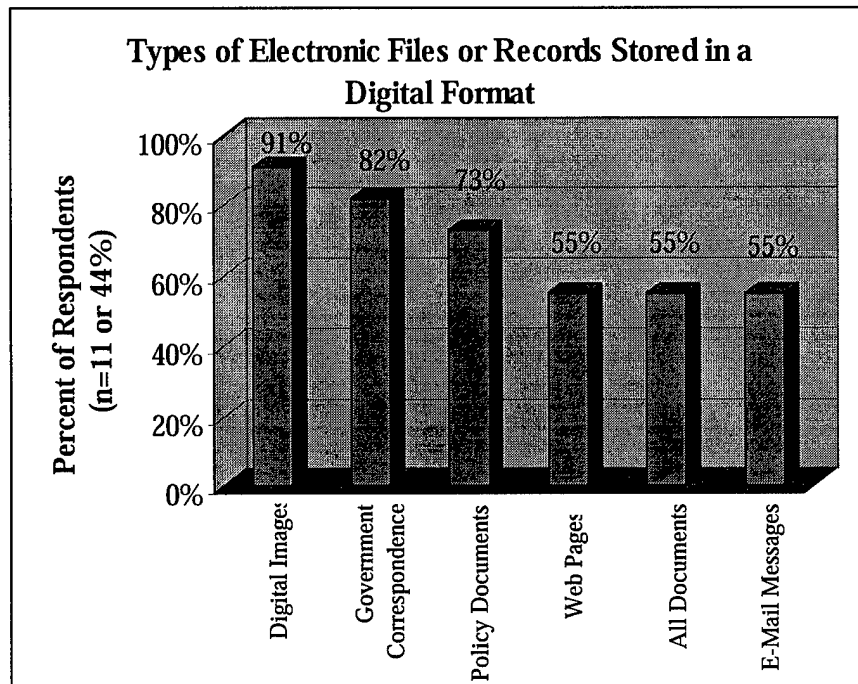
- a) 11 organizations responded that they maintain an electronic archive.
  - i) 82% of these organizations use Optical Disks.
  - ii) 73% of these organizations use Hard Disk Drives.
  - iii) 55% of these organizations use Magnetic Tape Backup.
  - iv) 36% of these organizations use CD-ROM Drives.



**FIGURE B3 – TYPE OF STORAGE MEDIA USED**

*Question 5: If your organization maintains an electronic archive, what types of electronic files or records are stored in a digital format?*

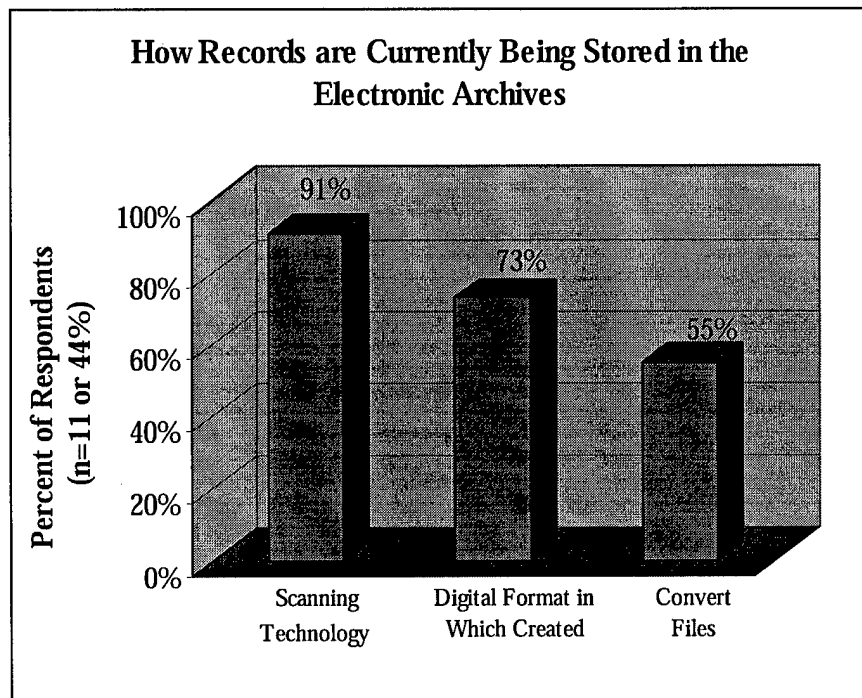
- a) 11 organizations responded that they maintain an electronic archive:
- i) 91% of these organizations stated that they store Digital Images.
  - ii) 82% of these organizations stated that they store Government Correspondence.
  - iii) 73% of these organizations stated that they store Policy Documents.
  - iv) 55% of these organizations stated that they store Web Pages.
  - v) 55% of these organizations stated that they store any document that is created within the organization.
  - vi) 55% of these organizations stated that they store E-mail messages.



**FIGURE B4 – TYPES OF ELECTRONIC FILES OR RECORDS STORED IN A DIGITAL FORMAT**

*Question 6: If your organization maintains an electronic archive, how are the records currently being stored in the electronic archives?*

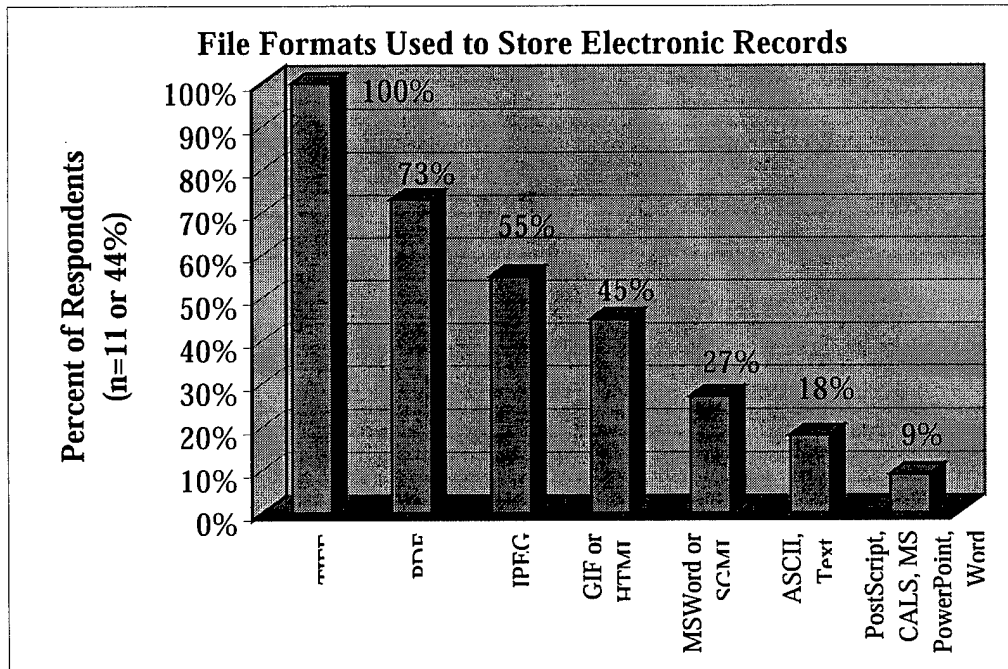
- a) 11 organizations responded that they maintain an electronic archive:
- i) 91% of these organizations use scanning technology used to create electronic images of paper documents.
  - ii) 73% of these organizations store electronic files in the digital format in which they were created.
  - iii) 55% of these organizations covert electronic files from their original format to a common standard format.



**FIGURE B5 – HOW RECORDS ARE CURRENTLY BEING STORED IN THE ELECTRONIC ARCHIVES**

*Question 7: If your organization maintains common file formats in an electronic archive, which of the following file formats are used to store the electronic records?*

- a) 11 organizations responded that they maintain an electronic archive:
- i) 100% of these organizations store TIFF formatted documents.
  - ii) 73% of these organizations store PDF formatted documents.
  - iii) 55% of these organizations store JPEG formatted documents.
  - iv) 45% of these organizations store GIF formatted documents.
  - v) 45% of these organizations store HTML formatted documents.
  - vi) 27% of these organizations store Microsoft Word (.doc) formatted documents.
  - vii) 27% of these organizations store SGML formatted documents.
  - viii) 18% of these organizations store text (.txt) formatted documents.
  - ix) 18% of these organizations store ASCII formatted documents.
  - x) 18% of these organizations store Excel spreadsheet (.xls) formatted documents.
  - xi) 9% of these organizations store PostScript formatted documents.
  - xii) 9% of these organizations store CALS formatted documents.
  - xiii) 9% of these organizations store Microsoft PowerPoint (.ptp) formatted documents.
  - xiv) 9% of these organizations store Word Perfect (.wpd) formatted documents.



**FIGURE B6 – FILE FORMATS USED TO STORE ELECTRONIC RECORDS**

*Question 8: If your organization maintains an electronic archive, what plans does the organization have for the records being stored?*

- a) 11 organizations provided a response to this question:
- i) 82% of the responding organizations plan on retaining records for 5 or more years.
  - ii) 55% of the responding organizations plan on destroying the records when no longer needed or required.
  - iii) 55% of the responding organizations plan to eventually transfer records to a storage facility or data warehouse.
  - iv) 27% of the responding organizations plan on replacing existing files with newer revisions of the same file.

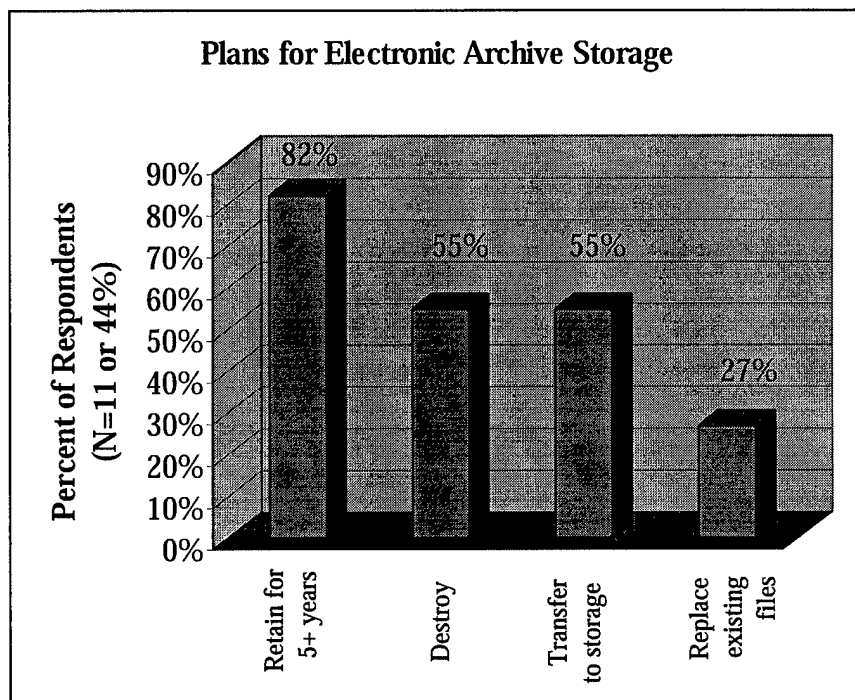


FIGURE B7 - PLANS FOR ELECTRONIC ARCHIVE STORAGE

*Question 9: If your organization maintains an electronic archive, what is the estimated current size of the electronic archive?*

- a) 10 organizations provided a response to this question:
- i) 70% of the responding organizations has an electronic archive with more than 10,000 records or files.
  - ii) 30% of the responding organizations has an electronic archive with less than 5,000 records or files.

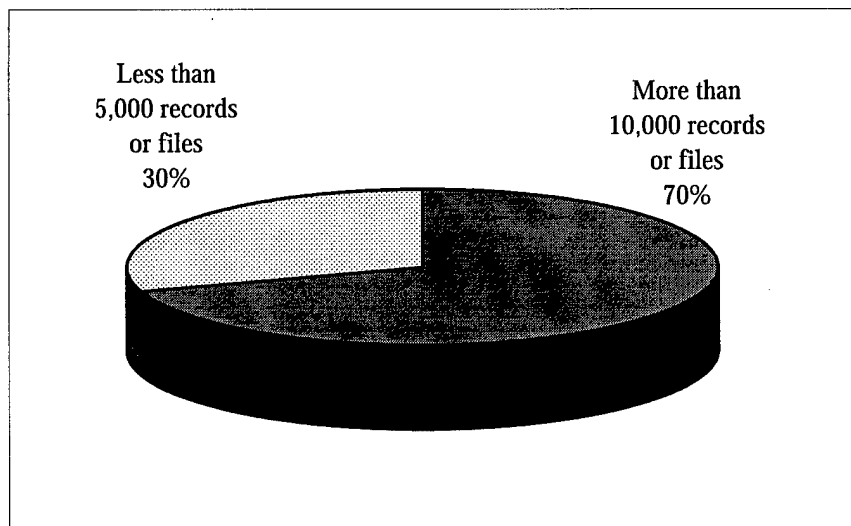
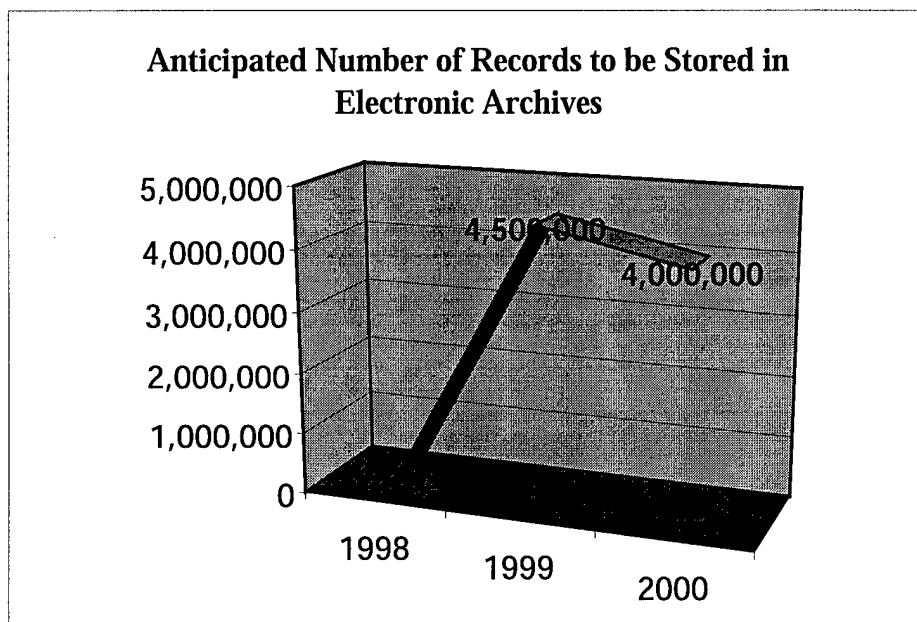


FIGURE B8 – CURRENT SIZE OF ELECTRONIC ARCHIVE

*Question 10: How many records do you anticipate to store in the electronic archive in each of the following years?*

- a) 7 organizations provided responses:
- i) For 1998: Range from 0 to 60,000 records.
  - ii) For 1999: Ranged from 3,000 to 4,500,000 records.
  - iii) For 2000: Ranged from 0 to 4,000,000 records.



**FIGURE B9 – ANTICIPATED NUMBER OF RECORDS TO BE STORED IN ELECTRONIC ARCHIVES**

*Question 11: Of the records identified in the question above, what percentage would you consider to be vital records or records that should be kept for historical purposes, and should be transferred to the National Archives or some other organization for permanent storage?*

- a) 13 organizations responded to this question:
- i) 54% of the respondents stated that **less than 10%** of the records should be permanently archived.
  - ii) 8% of the respondents stated that **40% - 50%** of the records should be permanently archived.
  - iii) 8% of the respondents stated that **50% - 60%** of the records should be permanently archived.
  - iv) 8% of the respondents stated that **70% - 80%** of the records should be permanently archived.
  - v) 8% of the respondents stated that **80% - 90%** of the records should be permanently archived.
  - vi) 15% of the respondents stated that **90% - 100%** of the records should be permanently archived.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

*Question 12: If your organization maintains an electronic archive, what is the estimated annual budget that the organization either spends or plans on spending in this effort?*

- a) 8 organizations provided responses to this question :

**TABLE B1 - ANNUAL BUDGET SPENT OR ANTICIPATED**

Year	#1	#2	#3	#4	#5	#6	#7	#8	Total
1998	\$2,000,000	\$800,000	\$200,000	\$140,000	\$100,000	\$ 0	\$ 0	\$100,000	\$ 3,340,000
1999	\$3,000,000	\$1,200,000	\$ 36,000	\$100,000	\$200,000	\$ 5,500,000	\$1,000,000	\$ 60,000	\$11,096,000
2000	\$4,000,000	\$ 0	\$ 36,000	\$ 90,000	\$100,000	\$ 5,500,000	\$2,000,000	\$ 60,000	\$ 9,786,000
2001	\$3,000,000	\$ 0	\$ 36,000	\$ 80,000	\$100,000	\$ 1,800,000	\$2,000,000	\$ 70,000	\$ 4,086,000
2002	\$2,000,000	\$ 0	\$ 36,000	\$ 50,000	\$300,000	\$ 1,800,000	\$2,000,000	\$ 70,000	\$ 4,256,000
Total	\$14,000,000	\$2,000,000	\$344,000	\$460,000	\$800,000	\$14,600,000	\$7,000,000	\$360,000	\$32,564,000

*Question 13: Does your organization currently use a Document Management System for the storage and retrieval of electronic files?*

- a) 50% (11 of 22) responding organizations **do** use a Document Management System.  
b) 50% (11 of 22) responding organizations **do not** use a Document Management System.

*Question 14: Which Document Management System software is in use? (9 Responses)*

DocsOpen [2]  
PC Docs  
FileNet  
Documentum  
Oracle  
Docupact  
KeyFile  
GOTS S/W being replaced by Quadra Star  
Highland's Higlview  
PRC's Productivity Edge  
Home-grown systems

*Question 15: Does your organization use a Records Management system?*

- a) 82% (18 of 22) responding organizations **do not** use a Records Management System  
b) 18% (4 of 22) responding organizations **do** use a Records Management System

*Question 16: If so, which Records Management System software is in use? (4 Responses)*

GOTS S/W being replaced by Quadra Star  
ForeMost 7.0  
Commercial application  
CHCS

*Question 17: If your organization does not currently maintain a document archive, are there plans to implement an archive in the near future?*

- a) 66% (4 of 6) respondents plan on implementing a document archive in the near future
- b) 33% (2 of 6) respondents **do not** plan on implementing a document archive in the near future

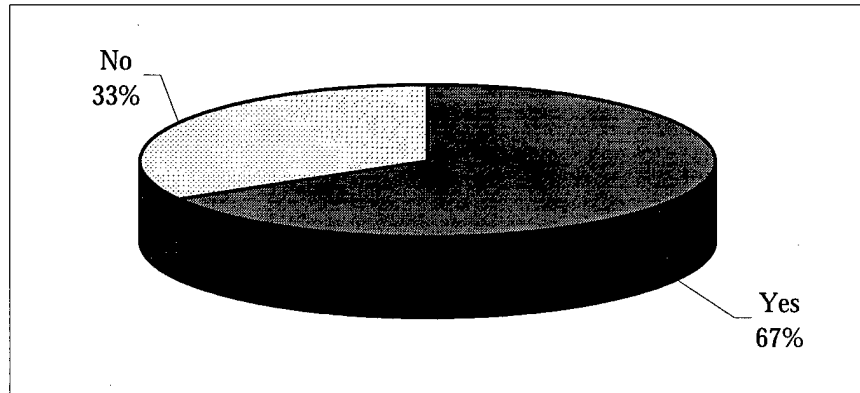


FIGURE B10 – PLANS TO IMPLEMENT AN ARCHIVE

*Question 18: Does your organization have plans for the future implementation of an electronic archive?*

- a) 88% (7 of 8) respondents **do** plan on implementing an electronic archive in the future.
- b) 12% (1 of 8) respondents **do not** plan on implementing an electronic archive in the future.

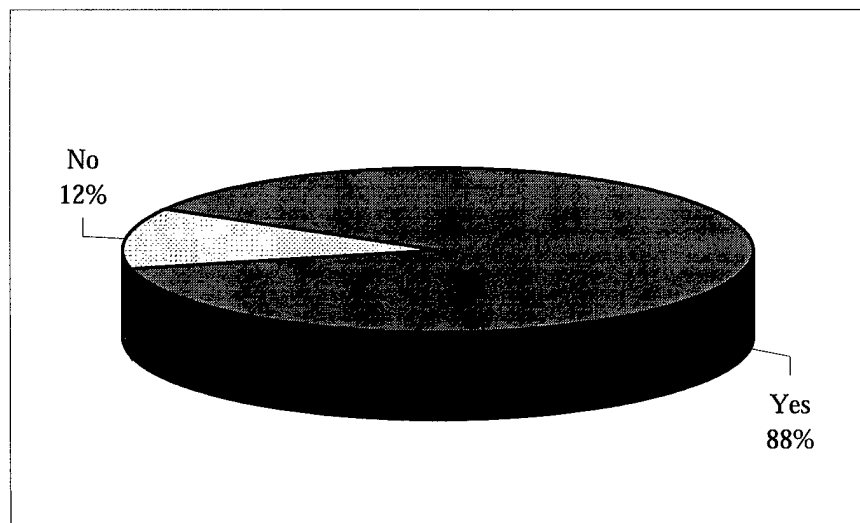


FIGURE B11 – PLANS FOR FUTURE IMPLEMENTATION OF AN ELECTRONIC ARCHIVE

*Question 19: Which of the following evolutionary paths do you foresee your organization pursuing in regards to storing digital information?*

- a) \*45% (10 of 22) of respondents are planning on implementing a Records Management System to maintain electronic files.
- b) \*41% (9 of 22) of respondents are planning on implementing a Document Management System to maintain electronic files.
- c) 32% (7 of 22) of respondents are committed to expanding or enhancing the existing electronic archives, such as moving in the direction of a full database implementation of electronic files that allows users to search for documents.
- d) 23% (5 of 22) of respondents are currently using and committed to maintaining an electronic archive of digital information.
- e) 18% (4 of 22) of respondents are either using or plan on using web technology to store documents or images in digital form.
- f) 5% (1 of 22) of respondents does not have any plans for storing digital information or electronic records.

\*Note: 4 respondents reported that they would be implementing both a Document Management System and a Records Management System.

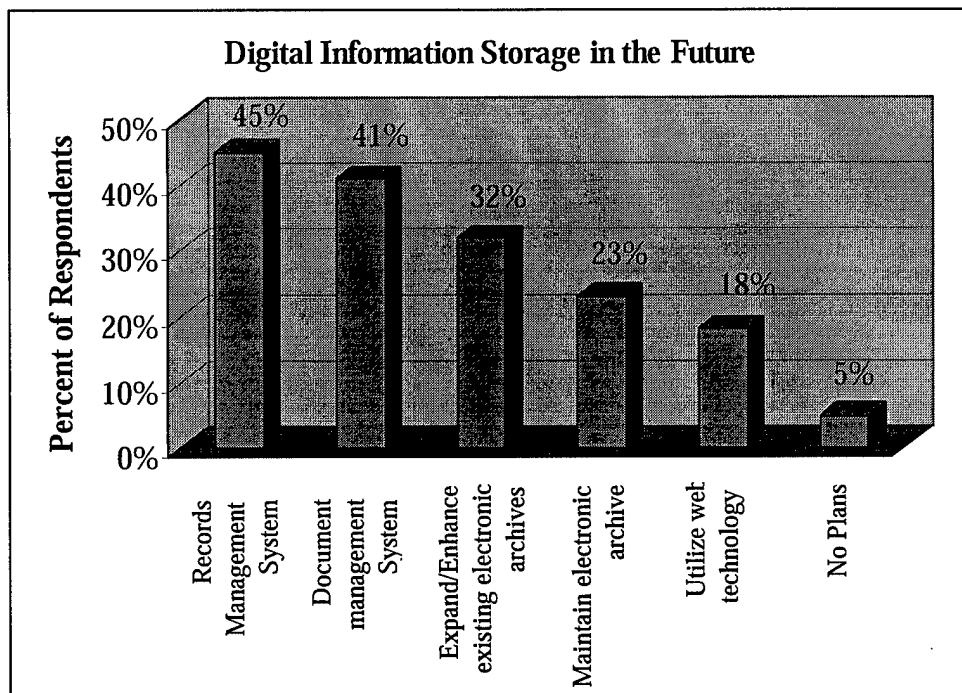
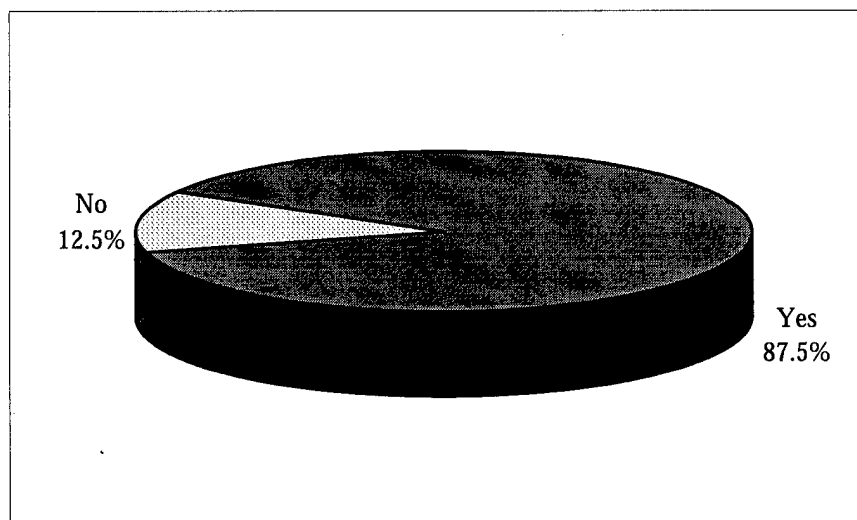


FIGURE B12 – DIGITAL INFORMATION STORAGE IN THE FUTURE

*Question 20: As an organization, would you like to be given direction or guidelines on how to establish, implement and maintain an electronic archive for digital records and images?*

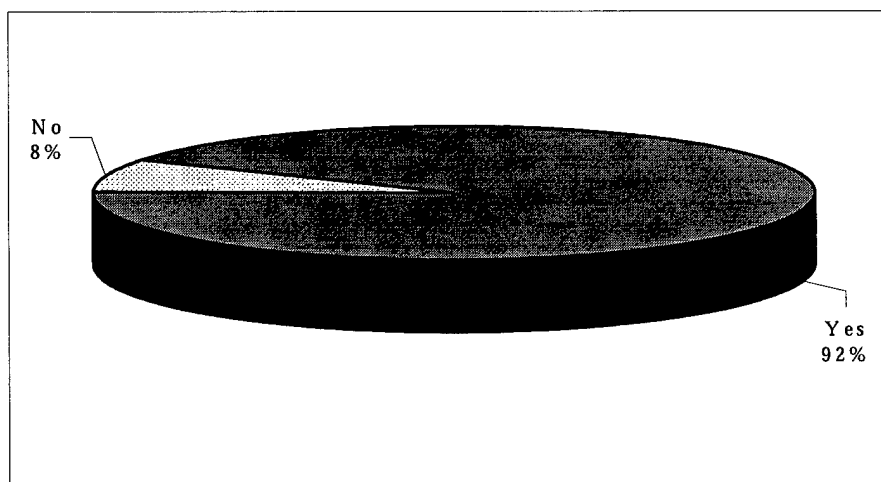
- a) 87.5% (21 of 24) respondents **would like** to receive direction.
- b) 12.5% (3 of 24) respondents **would not** like to receive direction.



**FIGURE B13 – WOULD LIKE TO BE GIVEN DIRECTION OR GUIDELINES ON HOW TO ESTABLISH, IMPLEMENT, AND MAINTAIN AN ELECTRONIC ARCHIVE FOR DIGITAL RECORDS AND IMAGES**

*Question 21: As an organization, do you feel that standards for how digital records should be stored in an electronic archive should be provided as guidelines by an organization such as the National Archives and Records Administration or the Office of the Secretary of Defense?*

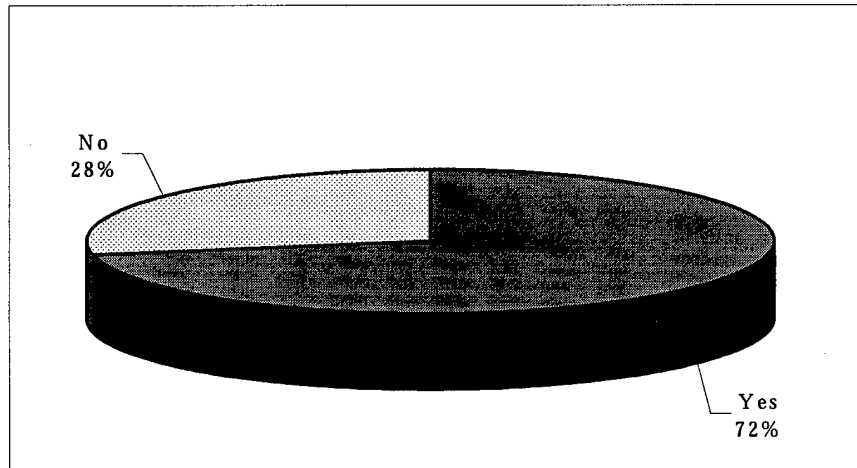
- a) 92% (22 of 24) respondents feel that standards **should** be provided.
- b) 8% (2 of 24) respondents feel that standards **should not** be provided.



**FIGURE B14 – SHOULD STANDARDS FOR DIGITAL RECORDS STORAGE IN AN ELECTRONIC ARCHIVE BE PROVIDED AS GUIDELINES BY NARA OR OSD?**

*Question 22: As an organization, do you feel that standards for how digital records should be stored in an electronic archive should be mandated by an organization such as the National Archives and Records Administration or the Office of the Secretary of Defense?*

- a) 72% (18 of 25) respondents feel that standards **should** be mandated.
- b) 28% (7 of 25) respondents feel that standards **should not** be mandated.



**FIGURE B15 – STANDARDS FOR HOW DIGITAL RECORDS SHOULD BE STORED IN AN ELECTRONIC ARCHIVE SHOULD BE MANDATED BY NARA OR OSD**

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

## IMAGING STANDARD SUPPORT TASK SURVEY

The primary purpose of this survey is to determine the level of effort DoD Government agencies are currently expending in the area of **digital imagery** and in the use of **electronic archives** to store **electronic files**. Other non-DoD Government agencies are also being asked to participate in this survey to determine the level of effort in this area outside of DoD

For this survey, the definition for a '**Digital Image**' is a computer (digital) representation of a picture. It may be a 'picture' of anything from a page of a document, a photograph, an x-ray, a map, a graph, etc.

An '**Electronic Archive**' can be a computer hard drive, a CD-ROM disk, a magnetic tape, an optical disk, or a floppy disk that is kept in a secure location for the purpose of maintaining a historical record of the information contained on the storage media. The usage of the term 'Electronic Archive' for this survey only applies to historical storage capabilities that are directly supported by the agency. The usage of the term 'Electronic Archive' does not apply to the efforts of individuals that determine on their own accord to maintain electronic files for their own reasons, whether it is for historical purposes or otherwise.

An '**Electronic File**' is any form of a document or image that is stored on electronic media, either as a computer file, a database, a web page, an email records file, etc.

*Use your TAB key or Mouse to move through the fields. For boxes, click on box to check, click again to un-check. For text field place cursor in shaded area and type, field will expand as needed.*

---

1. Does your agency currently store or archive documents for future use?

☐ Yes      ☐ No *[skip to Q.17]*

2. *[Answer if "yes" to Q1]* How are the documents archived? *[check one box]*

☐ Copies of paper documents stored in files or boxes  
☐ Electronic files stored on magnetic/optical media  
☐ Both

3. If your organization maintains a document archive, what is its purpose? *[check all that apply]*

☐ Allow information to be retrieved and shared throughout the organization  
☐ Retain records of documents for legal purposes  
☐ Retain records of documents for historical purposes  
☐ Required by federal, state or local laws

**[ANSWER QUESTIONS 4-12 ONLY IF YOUR ORGANIZATION CURRENTLY MAINTAINS AN ELECTRONIC ARCHIVE]**

4. If your organization maintains an electronic archive, what type of storage media is used? *[check all that apply]*

<input type="checkbox"/> Hard Disk	<input type="checkbox"/> Magnetic Tape Backup
<input type="checkbox"/> CD-ROM	<input type="checkbox"/> Optical Disk
<input type="checkbox"/> Unknown	<input type="checkbox"/> Other

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

5. If your organization maintains an electronic archive what types of electronic files or records are stored in a digital format? *[check all that apply]*

- |   |   |
|---|---|
| <input type="checkbox"/> Policy documents | <input type="checkbox"/> Government correspondence                    |
| <input type="checkbox"/> Email messages   | <input type="checkbox"/> Web pages                                    |
| <input type="checkbox"/> Digital images   | <input type="checkbox"/> Any document created within the organization |

6. If your organization maintains an electronic archive, how are the records currently being stored in the electronic archives? *[check all that apply]*

- ☐ Electronic files are stored in the digital format in which they were created  
☐ Scanning technology used to create electronic images of paper documents  
☐ Electronic files are converted from their original format to a common standard format such as HTML

7. If your organization maintains common file formats in an electronic archive, which of the following file formats are used to store the electronic records? *[check all that apply]*

- |  |                               |                               |                                    |                                   |
|--|-------------------------------|-------------------------------|------------------------------------|-----------------------------------|
| <input type="checkbox"/> JPEG                      | <input type="checkbox"/> PDF  | <input type="checkbox"/> STEP | <input type="checkbox"/> CADKey    | <input type="checkbox"/> SPIFF    |
| <input type="checkbox"/> CALS                      | <input type="checkbox"/> SGML | <input type="checkbox"/> HTML | <input type="checkbox"/> DWG       | <input type="checkbox"/> FlashPix |
| <input type="checkbox"/> GIF                       | <input type="checkbox"/> CGM  | <input type="checkbox"/> DXF  | <input type="checkbox"/> ME10      | <input type="checkbox"/> BIIF     |
| <input type="checkbox"/> TIFF                      | <input type="checkbox"/> IGES | <input type="checkbox"/> HPGL | <input type="checkbox"/> PostScrip | <input type="checkbox"/> VRML     |
| <input type="checkbox"/> Other(s) (please specify) |                               |                               |                                    |                                   |

8. If your organization maintains an electronic archive, what plans does the organization have for the records being stored? *[check all that apply]*

- ☐ Replace existing electronic files with newer revisions of the same file  
☐ Retain records for 5 or more years  
☐ Destroy records when no longer needed or required  
☐ Eventually transfer records to a storage facility or data warehouse

9. If your organization maintains an electronic archive, what is the estimated current size of the electronic archive?

- ☐ < 5,000 records or files  
☐ 5000 - 10,000 records or files  
☐ > 10,000 records or files

10. How many records do you anticipate to store in the electronic archive in
- |       |
|-------|
| 1998? |
| 1999? |
| 2000? |

11. Of the records identified in the question above, what percentage would you consider to be vital records or records that should be kept for historical purposes, and should be transferred to the National Archives or some other organization for permanent storage?

- |                                    |                                     |                                    |                                    |
|------------------------------------|-------------------------------------|------------------------------------|------------------------------------|
| <input type="checkbox"/> < 10%     | <input type="checkbox"/> 10% - 20%  | <input type="checkbox"/> 20% - 30% | <input type="checkbox"/> 30% - 40% |
| <input type="checkbox"/> 40% - 50% | <input type="checkbox"/> 50% - 60%  | <input type="checkbox"/> 60% - 70% | <input type="checkbox"/> 70% - 80% |
| <input type="checkbox"/> 80% - 90% | <input type="checkbox"/> 90% - 100% |                                    |                                    |

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

12. If your organization maintains an electronic archive, what is the estimated annual budget that the organization either spends or plans to spend in this effort?

1998?  
1999?  
2000?  
2001?  
2002?

13. Does your organization currently use a Document Management System for the storage and retrieval of electronic files? *[check one box]*

☐ Yes      ☐ No *[go to Q15]*

14. *[Answer if "yes" to Q13]* Which Document Management System software is in use?

15. Does your organization use a Records Management system?

☐ Yes      ☐ No *[go to Q19]*

16. *[Answer if "yes" to Q15]* If so, which Records Management System software is in use?

## ANSWER QUESTIONS 17 AND 18 ONLY IF YOUR ORGANIZATION DOES NOT MAINTAIN A DOCUMENT ARCHIVE

17. If your organization does not currently maintain a document archive, are there plans to implement an archive in the near future?

☐ Yes      ☐ No

18. Does your organization have plans for the future implementation of an electronic archive?

☐ Yes      ☐ No

## ALL RESPONDENTS PLEASE ANSWER QUESTIONS 19-22

19. Which of the following evolutionary paths do you foresee your organization pursuing in regards to storing digital information? *[Check one response]*

- ☐ The organization is currently using and committed to maintaining an electronic archive of digital information.
- ☐ The organization either uses or plans on using web technology to store documents or images in digital form.
- ☐ The organization is committed to expanding or enhancing the existing electronic archives, such as moving in the direction of a full database implementation of electronic files that allows users to search for documents.
- ☐ The organization is planning on implementing a Document Management System to maintain electronic files.
- ☐ The organization is planning on implementing a Records Management System to maintain electronic files.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

- ☐ The organization does not have any plans for storing digital information or electronic records.  
☐ OTHER (PLEASE SPECIFY)

20. As an organization, would you like to be given direction or guidelines on how to establish, implement and maintain an electronic archive for digital records and images? *[check one box]*

☐ Yes ☐ No

21. As an organization, do you feel that standards for how digital records should be stored in an electronic archive should be provided as guidelines by an organization such as the National Archives and Records Administration or the Office of the Secretary of Defense? *[check one box]*

☐ Yes ☐ No

22. As an organization, do you feel that standards for how digital records should be stored in an electronic archive should be mandated by an organization such as the National Archives and Records Administration or the Office of the Secretary of Defense? *[check one box]*

☐ Yes ☐ No

---

Name:  
Organization  
Address:

phone:  
Fax:  
e-mail:

Comments:

---

THANK YOU FOR TAKING TIME TO COMPLETE THIS SURVEY Once completed, please save your survey and *attach* to a new (or reply) message to:

[sue.h.mactavish@lmco.com](mailto:sue.h.mactavish@lmco.com)

Or fax to Sue MacTavish at (703) 671-3404.

## APPENDIX C - SURVEY RESPONDENTS

## Department of Defense

## Army Air Force Exchange Service

Jay McCartin  
3911 South Walton Walker Blvd.  
Dallas, TX 75236  
Phone: 214-312-2737  
FAX: 214-312-2896  
E-Mail: [mccartin@aafes.com](mailto:mccartin@aafes.com)

## Department of Defense

## Defense Commissary Agency

Sue W. Hall  
1300 E Avenue  
Fort Lee, VA 23801-1800  
Phone: 804-734-8817  
FAX: 804-734-8339  
E-Mail: [hallsw@hqlee.deca.mil](mailto:hallsw@hqlee.deca.mil)

## Department of Defense

## Defense Contract Audit Agency

Robert Wohlhueter  
8725 John J. Kingman Road - Suite 2135  
Fort Belvoir, VA 22060-6219  
Phone: 703-767-1036  
FAX: 703-767-1011  
E-Mail: [bwohlhueter@hq1.dcaa.mil](mailto:bwohlhueter@hq1.dcaa.mil)

## Department of Defense

## Defense Finance and Accounting Service

Pauline Korpanty  
1931 Jefferson Davis Highway  
Arlington, VA 22240-5291  
Phone: 703-607-3743  
FAX: 703-607-2773  
E-Mail: [pkorpanty@cleveland.dfas.mil](mailto:pkorpanty@cleveland.dfas.mil)

## Department of Defense

## Defense Information Systems Agency

Tommie Gregg  
DISA/CIO/IRM  
Virginia Square Plaza  
3701 North Fairfax Drive  
Arlington, VA 22203-1713  
Phone: 703-696-1890  
FAX: 703-696-1908  
E-Mail: [greggt@ncr.disa.mil](mailto:greggt@ncr.disa.mil)

## Department of Defense

## Defense Information Systems Agency

## Defense Information Technology Contracting Office

Cherie Cooper  
2300 East Drive - Bldg. 3600  
Scott AFB, IL 62225  
Phone: 229-9627  
FAX: 618-229-9683  
E-Mail: [cooperc@scott.disa.mil](mailto:cooperc@scott.disa.mil)

## Department of Defense

## Defense Logistics Agency

Allen Easterly  
8725 John J. Kingman Road - Suite 0119  
Fort Belvoir, VA 22060-6220  
Phone: 703-767-1135  
FAX: 703-767-5559  
E-Mail: [allen.easterly@hq.dla.mil](mailto:allen.easterly@hq.dla.mil)

## Department of Defense

## National Imagery and Mapping Agency

Russ Anderson  
NIMA - N-42  
1200 1<sup>st</sup> Street, SE  
Washington, DC 20303-0001  
Phone: 202-314-1056  
FAX: 202-314-1099  
E-Mail: [andersonr@nima.mil](mailto:andersonr@nima.mil)

## Department of Defense

## National Reconnaissance Office

George E. Darnell  
Information Management Group  
14675 Lee Road  
Chantilly, VA 20151-1715  
Phone: 703-808-5400  
FAX: 703-808-5082  
E-Mail: [darnellg@nro.mil](mailto:darnellg@nro.mil)

## Department of Defense

## National Security Agency

Susan A. Cook - S541  
9800 Savage Road - Suite 6886  
Fort Meade, MD 20755-6886  
Phone: 301-688-0094  
FAX: 301-688-2342  
E-Mail: None

## Department of Defense

## Office of Inspector General

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Raymond W. Braemer  
Forms & Publications Manager  
400 Army Navy Drive - Suite 408  
Arlington, VA 22202-2885  
Phone: 703-604-9781  
FAX: 703-604-9792  
E-Mail: [rbraemer@dodig.osd.mil](mailto:rbraemer@dodig.osd.mil)

**Department of Defense**  
**Office of the Secretary, JCS**  
Sterling S. Smith, Jr.  
Joint Staff/Joint Secretarial  
Information Management Division/RMAS  
BR  
Washington, DC 20318-0400  
Phone: 703-697-6906  
FAX: 703-695-7561  
E-Mail: [sterling.smith@js.pentagon.mil](mailto:sterling.smith@js.pentagon.mil)

**Department of Defense**  
**Department of the Army**  
**Records Management Division, TAG**  
Howard Greenhalgh  
6000 6<sup>th</sup> Street - Stop C-55  
Fort Belvoir, VA 22060  
Phone: 703-806-3258  
FAX: 703-806-3230  
E-Mail: [greenhalgh@rmipo.belvoir.army.mil](mailto:greenhalgh@rmipo.belvoir.army.mil)

**Department of Defense**  
**Department of the Navy**  
**Bureau of Medicine and Surgery**  
**Beneficiary Access and Support, 03**  
LT Mary Jenkins  
2300 E Street, NW  
Washington, DC 20372-5300  
Phone: 202-762-3143  
FAX: 202-762-3743  
E-Mail: [MEJenkins@us.med.navy.mil](mailto:MEJenkins@us.med.navy.mil)

**Department of Defense**  
**Department of the Navy**  
**Chief Information Officer**  
Charley Barth  
5312 Larochelle Drive  
Alexandria, VA 22315  
Phone: 703-602-6526  
FAX: 703-602-4668  
E-Mail: [barth.charley@hq.navy.mil](mailto:barth.charley@hq.navy.mil)

**Department of Defense**  
**Department of Health and Human Services**  
**Food and Drug Administration**

**Department of the Navy**  
**Office of the Assistant Secretary of the Navy**  
**(Financial Management and Controller)**  
Kathryn Bowman  
1000 Navy Pentagon  
Washington, DC 20350-1000  
Phone: 703-604-8249  
FAX: 703-604-6919/6921  
E-Mail: [bowman.kathryn@hq.navy.mil](mailto:b Bowman.kathryn@hq.navy.mil)

**Department of Defense**  
**Department of the Navy**  
**Judge Advocate General**  
CDR John V. Garaffa, JAGC, USN  
1322 Patterson Avenue, SE - Suite 3000  
Washington, DC 20374-5066  
Phone: 202-685-5295  
FAX: None  
E-Mail: [garaffajv@jag.navy.mil](mailto:garaffajv@jag.navy.mil)

**Department of Defense**  
**Department of the Navy**  
**Naval Criminal Investigative Service**  
Henry W. Persons, Jr.  
WNY Bldg 111 Attn: Code 27D  
716 Sicard Street, SE  
Washington, DC 20388-5380  
Phone: 202-433-9505  
FAX: 202-433-9518  
E-Mail: [hpersons@ncis.navy.mil](mailto:hpersons@ncis.navy.mil)

**Department of Defense**  
**Department of the Navy**  
**Naval Sea Systems Command**  
Bruce Maysmith  
SEALOG Atlantic, Code ND-051  
P.O. Box 100  
Indian Head, MD 20640-0100  
Phone: 301-743-6313  
FAX: 301-753-9526  
E-Mail: [maysmithbl@navsea.navy.mil](mailto:maysmithbl@navsea.navy.mil)

**Department of Defense**  
**Department of the Navy**  
**U.S. Marine Corps**  
Linda B. Goodwin  
Headquarters, U.S. Marine Corps  
2 Navy Annex, ARSE  
Washington, DC 20380-1775  
Phone: 703-614-1081  
FAX: 703-693-7270  
E-Mail: [lbgoodwin@notes.hqi.usmc.mil](mailto:lbgoodwin@notes.hqi.usmc.mil)

UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Seung Ja Sinatra  
FDA/OC/OIRM  
Room 16B-26  
56 Fishers Lane  
Rockville, MD 20857  
Phone: 301-827-4274  
FAX: 301-594-0060  
E-Mail: [ssinatra@bangate.fda.gov](mailto:ssinatra@bangate.fda.gov)

**Environmental Protection Agency**

Laura McHale  
401 M Street, SW (MC 3408)  
Washington, DC 20460  
Phone: 202-260-9329  
FAX: 202-401-8390  
E-Mail: [mchale.laura@epa.gov](mailto:mchale.laura@epa.gov)

**General Accounting Office**

Carol Hillier  
Records Officer  
441 G Street, NW - Room 7438  
Washington, DC 20016  
Phone: 202-512-4525  
FAX: 202-512-3366  
E-Mail: [hillierc.isc@gao.gov](mailto:hillierc.isc@gao.gov)

**Library of Congress**

Deborah Ramsey  
101 Independence Avenue, SE  
Washington, D.C.  
Phone: 202-707-6528  
FAX: 202-707-0633  
E-Mail: [dram@loc.gov](mailto:dram@loc.gov)

**Office of Personnel Management**

Mary Beth Smith-Toomey  
1900 E Street, NW - Room 5415  
Washington, DC 20415-7900  
Phone: 202-606-8358  
FAX: 202-418-3251  
E-Mail: [mbtoomey@opm.gov](mailto:mbtoomey@opm.gov)

**UNCLASSIFIED**

Imaging Standard Support Task

1 June 1999

(This space intentionally left blank.)

**UNCLASSIFIED**

## APPENDIX D – IMAGING STANDARD FOR ELECTRONIC RECORDS – ACTION PLAN

The following action plan was developed jointly by DoD and NARA with support from Lockheed Martin under the Imaging Standard Support Task Order. The goal of the plan was to identify a set of tasks and associated schedule to work with the other Federal Agencies, industry experts and academia to provide a solution for the DoD, Federal and NARA imagery standards archiving requirements.

The plan includes meetings with DoD, NARA, and other Government personnel to coordinate the activities related to scanned images of textual documents. Included with the scope of the study are consideration for access to this digitized archived material, the need to migrate to future technologies, a plan for comparative costs and advantages of particular COTS products, and the anticipated extent of use and volume throughout DoD.

The plan calls for Lockheed Martin to facilitate the development and presentation of a DoD-NARA sponsored conference designed to bring together subject matter experts from Government, industry, and academia on the subject of archival and access to electronic imagery records.

The following identifies the major tasks and schedule required toward providing a solution for the DoD, Federal and NARA imagery standards archiving requirements.

Task	Schedule
1. Update and expand information from the 1996-97 study on format/status options.	3 <sup>rd</sup> – 4 <sup>th</sup> quarter 1998
2. Initiate contacts with selected industry and academia representatives	3 <sup>rd</sup> – 4 <sup>th</sup> quarter 1998
3. Prepare and conduct a survey of DoD and other Federal Agencies to determine the level of effort Government agencies are currently expending in the area of digital imagery and in the use of electronic archives to store electronic files.	4 <sup>th</sup> quarter 1998
4. Publish preliminary findings on imaging format(s) selections with associated cost, and migration data supporting their selection.	1 <sup>st</sup> quarter 1999
5. Hold Invitational Conference for Federal, Industry and Academia representatives	2 <sup>nd</sup> quarter 1999
6. Publish recommendations and findings on imagery standards for electronic records	2 <sup>nd</sup> quarter 1999
7. Develop archival guidelines.	3 <sup>rd</sup> quarter 1999
8. Collect and revise NARA guidelines and disseminate for community review.	4 <sup>th</sup> quarter 1999
9. Propose Title 36 CFR modification to reflect new guidelines for archiving of electrical imagery records.	1 <sup>st</sup> quarter 2000

(This space intentionally left blank.)

## APPENDIX E – DoD-NARA CONFERENCE

Over the last several years, the Department of Defense (DoD) and the National Archives and Records Administration (NARA) have sponsored a series of studies and conferences on the topic of digital image standards and the selection of the most appropriate digital imaging standard for long-term preservation of Federal documents. The most recent, The DoD-NARA Scanned Images Standards Conference, held March 31- April 1, 1999, was attended by over 90 individuals eager for an opportunity to learn and exchange information on the current status of imaging in DoD and other Federal agencies. The Program included an overview of imaging standards including the types and extent of their use, and the status of selected imaging projects and standards associated with imaging.

The first day of the conference was focused on a looking at "Why we are here" and "What's Going On." Welcoming comments were made by Dr. Ken Thibodeau, Director, Electronic Records Programs at National Archives, and Burt Newlin, from the DoD, OSD C3I, the conference sponsors. Then Sue MacTavish of Lockheed Martin, the Conference facilitator and Project Manager for the DoD Imaging Standards Policy Support task, kicked off the conference with a presentation on "Why are we here?" Because:

- Electronic Records have become a very HOT topic.
- The use of computers is changing the way government documents are created, accessed and managed. Electronic records, the Internet and E-mail have become an increasingly large part of the everyday work environment. To improve access, distribution, and interoperability, Federal agencies are converting large numbers of documents from paper to electronic digital images. Increased accessibility to the most current data drives the move away from paper records whenever possible. Among these Federal agencies there is increasing interest in receiving National Archives and Records Administration (NARA) guidance identifying acceptable digital image formats for long term preservation.
- For Federal records requiring permanent retention, long-term preservation of digitally imaged records has become problematic. While the advantages of digitally imaged documents are tremendous, due to the relatively short life cycle of digital image technology (both hardware and software), it is commonly accepted that all formats used today will eventually become obsolete.
- Computer tapes and disks deteriorate, and the hardware and software systems on which they can be read become obsolete. For an electronic record long term preservation requires that as the technology changes that the record be migrated from one format to another and then verified to ensure no loss of data. Limiting the number of image formats to monitor for technology change becomes an essential part of long-term preservation strategy. Identification of appropriate and relatively stable formats is key to success.

Sue's introductory remarks were followed by speakers for various Agencies providing an update on "What's going on"

- Dr. Scott Lackey from the Center for Army Lessons Learned (CALL) reported on how CALL is moving away from acquiring paper and encouraging pure electronic record acquisitions, because they provide more utility to end-users. CALL utilizes the DoD 5015.2-STD metadata requirements as the basis for their system. The metadata is linked to the actual electronic or converted record. With all record components managed as one record.
- Ms. Bette Mahoney from the Defense Human Resources Activity (DHRA), Joint Requirements and Integration Office (JR&IO), briefed the group on the Defense Personnel Records Imaging System (DPRIS). DPRIS is an OSD initiative to move toward a common operating environment for electronically querying Official Military Personnel File (OMPF) records systems. All of the Services have converted or are converting their personnel records to digital images in the TIFF format, but have not utilized common indexing and system architecture. Therefore, while all these records are in TIFF format there are dissimilar header structures and use of TIF extensions. DPRIS employs Web technologies to support electronic queries of these disparate OMPF systems and speed up search

response times. OMPF records are by far the most used and therefore the most expensive retired DoD records to support. They are essentially archived "forever" and there is a need for identical standards for active, retired and archived records.

- Updating the group on the status of the government wide document declassification program was Kirk Lubbes, President of Records Engineering, LLC. It is estimated that in the CIA alone there are some 40 million pages of information scheduled to be reviewed for declassification. That would equal 40 Washington Monuments in height if they were stacked. The documents are scanned in TIFF 6, Group 4 compression. Each document is indexed with up to 26 bibliographic fields. The Electronic Document Interchange Standard (EDIS), a voluntary standard for electronic document interchange among Executive Branch agencies, was developed for the declassification project. The standard governs both document metadata and document images that are to be exchanged for purposes of coordinate review, as well as minimum transfer metadata. Digitizing these records facilitates declassification efforts and FOIA requests.
- Steve Wehrly of the U.S. Army Publishing Agency provided an overview of electronic information publishing activities in the Army. His presentation covered the myriad of Army publishing media for: Administrative publications (including directives); Training and Doctrine; and Technical and Equipment. He reviewed the history of Army Electronic Publishing; their use of the Web; and the Army's "Less-paper" Policy. Steve concluded with a discussion on linear and non-linear media; electronic publications and interactive electronic publications; and the issue surrounding archiving of these electronic publications and interactive electronic publications.
- Steve Puglia of NARA presented information on the findings and guidelines of the Electronic Access Project at NARA their findings and guidelines. Steve's data illustrated clearly the fact that it is not the longevity of digital optical media (30 to 200 years), but rather the digital data system's 5 to 10 year systems life that is the critical factor in migrating data to new technology or system. Leading to the conclusion that electronic imaging is excellent for access and rapid retrieval, but lousy for long-term preservation.

Two more information sessions followed the "What's going on" presentations:

- George Wenchel, of Lockheed Martin, provided a basic overview of the myriad of standards available, and a discussion of the pros and cons of de facto versus de jure standards. George focused the majority of his comments on TIFF, PDF, and BIIF (the new ISO standard 12087-5, 1998). He emphasized that while there are currently no digital image formats that are acceptable for long-term preservation, the goal is to identify formats that are likely to live longer than others in guidelines as approved data preservation formats. By selecting such standards, NARA will be able to reduce the frequency of data reformatting required to migrate data through different standards and technology and thus to minimize the cost of digital image data preservation.
- Mike Pickard, also of Lockheed Martin, presented data collected in survey of DoD and selected other Federal Agency records managers re activities and plans in the area of electronic records management. The Imaging Standard Support Task Survey was sent to 35 Federal Records Officers in the DoD and selected Federal Agencies in October 1998. Results were collected from 25 Agencies – a 71% return ratio. The purpose of the survey was to help determine the current level of effort DoD Agencies were expending in archiving electronic records and to determine which electronic formats were currently being used to generate digital images. The top six responses to this latter inquiry were either de facto standards, proprietary file formats, or 'unofficial' forms of approved standards (HTML is a form of SGML). TIFF was used by 100% and PDF by 73% of the respondents that were using digitally imaged documents.

The second day of the conference was devoted to small group discussions and idea generation.

The groups were asked to discuss what are the drivers, and roadblocks to a successful digital imaging program, and who should be doing what, when and how. The key drivers seem to be access and FOIA. The key roadblock was costs and lack of management understanding of the need for appropriate funding in this area. The lack of standards

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

was not seen as a major roadblock. The group felt strongly that a united government voice was needed, with strong NARA leadership and a means of sharing data.

In the afternoon the groups were asked to review, and comment on the recommendations found in the preliminary study report which was disseminated to conference attendees. These recommendations, as found in the Preliminary Report, are:

- Image materials in the most stable, uncompressed format available.
  - For Personnel records: Image using TIFF for archiving, TIFF or PDF formats for access. Convert all Current TIFF 4 images to one standardized format.
  - For Declassified Records: Image using TIFF for archiving, TIFF 6 or PDF formats for access. Historically significant records should be converted to paper, microfilm or ASCII formats.
  - For manuals, standards, directive type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML or SGML formats for dissemination.
- Develop standard header data guidelines for the TIFF image format.
- Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.
- Establish criteria for selection of digital images for accessioning in the National Archives.
- Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.
- Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.
- Plan for migration of digital images every 3-5 years with of cost equivalent to 50 – 100% of the costs associated with original imaging project.
- Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
- Develop standard set of metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set.
- Establish guideline describing metadata that must accompany digital image when submitted for archival accessioning.
- Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
- Convert documents that require long-term preservation from application format to an image format for storage.

Most thought that the recommendations should be grouped, but there was no consensus to eliminate any single recommendation nor was there any clear consensus on what recommendations were the most important. One of the small groups proposed an implementation approach:

1. Manage the process (records management, management and policy).
2. Study, plan, gather information through cost/benefit analysis of entire life-cycle (especially document preparation, searching, and migration).
3. Pick an interim standard during step 2, which will be accepted and supported by DoD and NARA – this will enable the cost-benefit analysis to be conducted.
4. Practice migration and preservation while documents are in active use.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

This conference was conducted as part of the Office of the Secretary of Defense/Command, Control, Communications, and Intelligence (OSD/C3I) sponsored Imaging Standard Support Task Order and was facilitated by Sue MacTavish of the Lockheed Martin Imaging Standard Support Team. The data collected at the conference will be folded in to the study's final report which will be available the end of May.

## Drivers

### Green X

#### Legal/Law

- Executive orders
- Directives
- FOIA

- What the public expects
- Government re-invention
- Compatibility between agencies
- Industry provided solution
- COTS

### Green Dot

- Access – internal and external to the agency
- EFOIA

- Legal – Keep and compatibility with legal system

- Cost Savings

  - Less Paper
  - Less Storage

- Regulation

- Items created electronically

### Blue Dot

#### Access

- Searchability, public access
- Seamless access
- Historical documents – multiple access

#### Mandate

- Legal requirement for access (FOIA)
- Paperless workplace

#### Records Management

- Paperless process
- Source of information is electronic
- E-Mail
- Word processing, spreadsheets

### Blue X

#### Title 44

#### Mission Accomplishment

- Need Information
- Access – speed
- Preserve

#### Paper Records

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Labor intensive

Costly – Resource reduction period

Space considerations

Paperwork Reduction Act

EFOIA

## Red X

Volume growing – Staff declining

“Solutions” from vendors & Policy types

Access – within agency, by public, WWW

Growing expectations for info, delivery

Associations – AIIM ARMA, NAGARA

Centralized source purchases

## Red Dot

Better in-house and public access

Being able to accept electronic records for evidentiary purposes

Doing more with less

Cost savings

Printing

Storage

Distribution

Compliance with present and existing regulations

## Roadblocks

## Blue Dot

Funding

Lack of

Limited

Added tasks

Knowledge

Lack of knowledge on problems inherent with imaging

Training issues

Management Resistance

Duplication of effort

Lack of information being shared (experience)

Technology

Constantly changing

Uncertainty of formats for future purposes

Direction/Guidance

No best practices

Lack of advisory council

Lack of standards

## Blue X

Federal Agency Records Management

Dysfunctional

Legal Acceptance

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

- Lawyers
- Unfunded Requirements
  - Falls out of POM process
- Speed/Volume of Generation of Information
- Pace of Technology Change
- Lack of Senior Management understanding of records requirement

## Red Dot

- Resources
  - Personnel, Money & Equipment
- Lack of Standards/Guidance
- Incompatible Formats
- Poor Records Management Programs

## Red X

- CIO Council
  - Lacks broad vision
  - Not leading
  - IT focus
- Agencies
  - Culture, comfort level
  - Budget
    - Cycle, new costs, procurement, competition
    - Full Costs (maintenance, migration)
- Marketplace
  - No one path,
  - Competition
  - False claims
  - No interconnectivity
- Standards
  - None
  - Too slow to change
  - Vendor application
- Preservation/Archiving
  - Legal
    - Issues of evidence, signature, approval, image certification
  - No leadership from NARA, DoD, CIO, DoJ
  - Lack of Knowledge
    - Best practice, cost effective, past experience
  - Technology
    - Help or Hinder
  - Non compliance with standards, schedules
  - Perception
    - Do what is trendy

## Green X

- Technology Solution -- Share Responsibility
- Vs.
- Management Solution
- Combination of Electronic and paper records

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

Cost  
Security  
Signatures

## Green Dot

Costs  
    Budget Cycles  
    Management understanding, lack of  
Technology change  
    Vendors  
CIOs  
Lack of clear guidance/direction  
Lack of Standards  
Legal acceptance  
    Signature

## Who What When

## Red Dot

Agency top Management  
    React to demand  
Agency Management  
    Encourage change  
NARA & OSD  
    Continue/complete process of providing guidance  
Vendor  
    Need to comply  
OMB  
    Raise consciousness of congressional oversight  
Agency  
    Records management offices need to be synchronized  
Federal employees  
    Need to know and follow regulations  
NARA with agencies *i.e.* NOAA, NASA, -  
    Resolve technical issues  
    Establish standards  
    Share info with private sector  
NARA  
    Have resources to work with agencies  
Set standards & Technical compatability  
Cooperate/communicate  
    Public & private sectors  
Support each other

## Red X

NARA  
    Start in 1983  
    Leadership  
        Best practices, clearinghouse, agencies, vendors, associations, states  
    Firm

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

## Agencies

- Slower, more care, mid-curve, share
- Interoperability Follow standards
- Educate, share
- Consolidated buying

## Legal

- Courts & DoJ
- Evidence, precedence, guidance

## GSA

- Procurement regulations

## CIO

- Refocus, educate
- Best cases
- Partner with industry
- Incentives/penalties

## Vendors

- Accountability
- Interoperability
- Deliver what promised, when
- Longer life-cycle support

## Standards

- Quicker, flexible, reflect market

## Green X

### Unified Government voice

- Standards *i.e.*
- Follow-on to this conference

Drive vendors to open formats

Develop open formats/standards

Develop roadmaps for agencies

Industry solution-

- What is it?

- Conference Microsoft and Adobe

Address records management solutions

## Green Dot

### United Government voice

- NARA leadership
- Information exchange
- Best Practices

Agency top management – understand issue

- Funding

- Training

Vendor compliance with “united” Government

- Move vendors towards open formats

Advisory Council

Clearing house

## Blue Dot

Establish Advisory Council

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NARA advise agencies on standards/processes  
Establish Dept with in agency:  
Work with NARA on best process for agency  
Disseminate information within agency  
Govern implementation of policy  
CIO/Electronic records management  
Greater Government attention to funding need of electronic records

## Blue X

Records are kept for agency purposes  
No target format  
What is the format for archiving?  
NARA – not Congress needs to decide – lead the process

## Recommendations

### Blue Dot

1. Image materials in the most stable, uncompressed format available.
    - For Personnel records: Image using TIFF for archiving, TIFF or PDF formats for access. Convert all Current TIFF 4 images to one standardized format.
    - For Declassified Records: Image using TIFF for archiving, TIFF 6 or PDF formats for access. Historically significant records should be converted to paper, microfilm or ASCII formats.
    - For manuals, standards, directive type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML or SGML formats for dissemination.
  1. Convert documents that require long-term preservation from application format to an image format for storage.
  1. Establish criteria for selection of digital images for accessioning in the National Archives.
  2. Develop standard header data guidelines for the TIFF image format.
  2. Develop standard set of metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set and tag as MARC records.
  2. Establish a guideline describing metadata that must accompany digital image when submitted for archival accessioning.
  3. Plan for migration of digital images every 3-5 years with of cost equivalent to 50 - 100% of the costs associated with original imaging project.
  3. Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
  4. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.
  4. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
    - Low. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.
- ~~Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.~~

## Follow On

### Blue Dot

1. Establish NARA acceptable standard for images (urgent need)  
Environmental storage issues
1. Develop standard GUI metadata front end for imaging when scanning – include color bar, resolution  
Standards should evolve with time
2. Keep digital masters off line

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

- manipulate on line copies
- Manage image life cycle migration
- Develop standard operating procedures and documentation to recover costs for providing access
- Find better ways to fund projects
- Improve project planning
- 4. National Digital Library - Best Practice
  - Establish test beds providing positive examples, identify issues and pit falls
- 3. Acquire empirical data/metrics – for managerial purposes
  - Creation of a certified Digital Archive who will manage images
  - Internal/external organizational structure to support those efforts
- 5. Insure that records schedules address electronic records

## Recommendations

### Blue X

1. Establish archival standards for pure electronic records and electronic images that meet legal requirements
1. Image materials in the most stable, uncompressed format available.
  - For Personnel records: Image using TIFF for archiving, TIFF or PDF formats for access. Convert all Current TIFF 4 images to one standardized format.
  - For Declassified Records: Image using TIFF for archiving, TIFF 6 or PDF formats for access. Historically significant records should be converted to paper, microfilm or ASCII formats.
  - For manuals, standards, directive type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML or SGML formats for dissemination.
1. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.
2. Use DoD 5015.2 STD for electronic storage
3. Plan for migration of electronic records every 3-5 years with of cost equivalent to 50 - 100% of the costs associated with original imaging project.
4. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.

~~Develop standard header data guidelines for the TIFF image format.~~

~~Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.~~

~~Establish criteria for selection of digital images for accessioning in the National Archives.~~

~~Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.~~

~~Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.~~

~~Develop standard set of metadata of textual digital images using DoD 5015.2 STD, EAD, and Dublin Core as minimum set and tag as MARC records.~~

~~Establish guideline describing metadata that must accompany digital image when submitted for archival accessioning.~~

~~Convert documents that require long term preservation from application format to an image format for storage.~~

## Follow On

### Blue X

1. Use DoD 5015.2 Std for metadata
2. Study and evaluate migration strategies to reduce impact of 3-5 year upgrade cycle
3. Study and evaluate formats for non-textual material (photos, X-rays, maps, etc....)
4. Image materials in the most stable format available
5. Study and evaluate De Jure formats

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

6. Budget for migration to long term archiving format
7. NARA needs to lead/organize study groups to develop standards for imaging and electronic storage
8. DoD needs to work to develop compatibility within DoD

## Recommendations

### Green X

1. Image materials in the most stable, uncompressed format available.
  - For Personnel records: Image using TIFF for archiving, TIFF or PDF formats for access. Convert all Current TIFF 4 images to one standardized format.
  - For Declassified Records: Image using TIFF for archiving, TIFF 6 or PDF formats for access. Historically significant records should be converted to paper, microfilm or ASCII formats.
  - For manuals, standards, directive type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML or SGML formats for dissemination.
1. Develop standard header data guidelines for the TIFF image format.
1. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.
1. Establish criteria for selection of digital images for accessioning in the National Archives.
1. Develop standard set of metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set and tag as MARC records.
1. Establish a guideline describing metadata that must accompany digital image when submitted for archival accessioning.
1. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
1. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.
2. Plan for migration of digital images every 3-5 years with of cost equivalent to 50 - 100% of the costs associated with original imaging project.
2. Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
3. Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.
3. Convert documents that require long-term preservation from application format to an image format for storage.

## Additions

### Green X

Add from group sessions

CIO's

Meeting Ideas - Follow on conferences

Add PDF

Priority:

Provide Standards

Format - TIFF

Scanning practices

Metadata

## Follow On Conference

### Green X

1. DoD/Government Agencies

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

2. Industry
3. Library of Congress
4. Standards bodies i.e. ISO/AIIM/ANSI
5. ARMA
6. State Governments
7. CIO's - Government
8. Universities
9. Independent bodies

## Recommendations

### Green Dot

1. Image materials in the most stable, uncompressed format available.
  - For Personnel records: Image using TIFF for archiving, TIFF or PDF formats for access. Convert all Current TIFF 4 images to one standardized format.
  - For Declassified Records: Image using TIFF for archiving, TIFF 6 or PDF formats for access. Historically significant records should be converted to paper, microfilm or ASCII formats.
  - For manuals, standards, directive type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML or SGML formats for dissemination.
2. Develop standard header data guidelines for the TIFF image format. (Technical and Contextual)
3. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.
4. Establish criteria for selection of digital images for accessioning in the National Archives.
5. Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.
6. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.
7. Plan for migration of digital images every 3-5 years with of cost equivalent to 50 - 100% of the costs associated with original imaging project.
8. Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
9. Develop standard set of metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set and tag as MARC records.
10. Establish a guideline describing metadata that must accompany digital image when submitted for archival accessioning.
11. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
12. Convert documents that require long-term preservation from application format to an image format for storage. Text as Text and Images to application independent format.
13. Digital fingerprinting or validation to prove images are unaltered

## Follow On

### Green Dot

Records management is a priority.

- Plan for migration (7)
- Study and evaluate De Jure interchange formats (6)
  - Study and evaluate migration strategies (8)

File Format (1)

Header Data (2&3)

Indexing and Metadata (9&10)

Criteria of selection for accessioning are based on record information, not the format. (4)

**Recommendations**Red Dot

1. Develop standard header data guidelines for the TIFF image format.
2. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data. Image materials in the most stable, uncompressed format available.
3. Image materials in the most stable, uncompressed format available.
4. For Personnel records: Image using TIFF for archiving, TIFF or PDF formats for access. Convert all Current TIFF 4 images to one standardized format.
5. For Declassified Records: Image using TIFF for archiving, TIFF 6 or PDF formats for access. Historically significant records should be converted to paper, microfilm or ASCII formats.
6. For manuals, standards, directive type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML or SGML formats for dissemination.
7. Plan for migration of digital images every 3-5 years with of cost equivalent to 50 - 100% of the costs associated with original imaging project.
8. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.
9. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
10. Establish a guideline describing metadata that must accompany digital image when submitted for archival accessioning.  
 Establish criteria for selection of digital images for accessioning in the National Archives.  
 Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.  
 Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.  
 Develop standard set of metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set and tag as MARC records.  
 Convert documents that require long-term preservation from application format to an image format for storage.

**Recommendations/Priorities**Red X

1. Study and evaluate de jure interchange formats for long-term archive acceptance and application in the field.
2. Plan for migration of digital images every 3-5 years with of cost equivalent to 50 - 100% of the costs associated with original imaging project.
3. Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
4. Develop standard header data guidelines for the TIFF image format.
5. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize TIFF header data.
6. Image materials in the most stable, uncompressed format available.
  - a. For Personnel records: Image using TIFF for archiving, TIFF or PDF formats for access. Convert all Current TIFF 4 images to one standardized format.
  - b. For Declassified Records: Image using TIFF for archiving, TIFF 6 or PDF formats for access. Historically significant records should be converted to paper, microfilm or ASCII formats.
  - c. For manuals, standards, directive type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML or SGML formats for dissemination.
7. Develop standard set of metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as minimum set and tag as MARC records.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

8. Establish criteria for selection of digital images for accessioning in the National Archives.
9. Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.
10. Establish a guideline describing metadata that must accompany digital image when submitted for archival accessioning.
11. Study and evaluate formats designed for non-textual material, e.g. photography, aerial imagery, x-rays, radar, for compatibility with textual digital image formats in the archive environment.
12. Convert documents that require long-term preservation from application format to an image format for storage.
13. Digital fingerprint or validation to prove that it is unaltered.

## Grouped

1. Manage the process
  - a. Records Management
  - b. Management/Policy
  - c. Technology is not the issue
2. "Study, Plan and Gather Information"
  - a. Cost-Benefit analysis of entire Life cycle
    - i. ESP Doc Prep, searching, migration
  - b. Publicize cost studies, best practices
3. Standards
  - a. Interim Standards during "2."
    - i. Accepted, supported by DoD & NARA
  - b. Permanent standard(s)
4. Migration/Preservation
  - a. In Active use
  - b. Archive

## Additional Recommendations

1. Metadata
  - a. Establish guidelines describing metadata that must accompany digital images when submitted for archival accessioning.
  - b. Develop standard header data guidelines for the TIFF image format.
  - c. Develop standard sets of metadata of textual digital images using DoD 5015.2-STD, EAD, and Dublin Core as a minimum
2. Image Formats
  - a. Image materials in the most stable format available.
    - i. For personnel records: Image using TIFF for archiving or PDF formats for access. Convert all current TIFF 4 images to one for standardized formats.
    - ii. For declassified records: Image using TIFF for archiving, TIFF6 or PDF formats for access. Historically significant records should be converted to paper, microfilm, or ACSII formats.
    - iii. For manuals, standards, directives type material: Image using TIFF, ASCII and ASCII SGML tagged files for archiving. PDF, HTML, or SGML formats for dissemination.
3. Standards
  - a. Work with Association for Information and Image Management (AIIM) and American National Standards Institute (ANSI) to standardize the TIFF header data.
    - i. Develop standard header data guidelines for the TIFF image format.
  - b. Work with NARA.
    - i. Establish criteria for selection of digital images for accessioning into the National Archives.
    - ii. Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.

- iii. Establish guidelines describing metadata that must accompany digital images when submitted for archival accessioning.
  - iv. Convert documents that require long-term preservation from application format to an image format for storage.
  - c. Study and evaluate *de jure* interchange formats for long-term archive acceptance and application in the field.
4. Preservation Migration
- a. Accession digital images that have been imaged in the most stable format available and those that meet the selection criteria.
  - b. Plan for migration of digital images every 3-5 years with costs equivalent to 50-100% of the costs associated with the original imaging process.
  - c. Study and evaluate migration strategies applied to digital data archives to application in the maintenance of textual digital images.
  - d. Convert documents that require long-term preservation from application format to an image format for storage.
  - e. Study and evaluate formats designed for non-textual materials, e.g. photography, aerial images, x-rays, radar, for compatibility with textual digital image formats in the archive environment

(This space intentionally left blank.)

UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

## APPENDIX F - CONFERENCE ATTENDEES

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Hixson, Larry	Administrative Office of the U.S. Courts One Columbus Circle, NE OIT-TEO, Suite 3-560 Washington, DC 20544	202-502-2762	202-502-2766	larry_hixson@ao.uscourts.gov
Lagana, Joyce	Administrative Office of the U.S. Courts One Columbus Circle, NE Washington, DC 20544	202-502-2331	202-502-2366	joyce_lagana@ao.uscourts.gov
Levenson, Stephen	Administrative Office of the U.S. Courts 6194 Wild Valley Court Alexandria, VA 22310	202-502-2625		levenson@ao.uscourts.gov
Bousquet, Dorothy	Army and Air Force Exchange Service ATTN: FA-C P.O. Box 660659 Dallas, TX 75266	214-312-6121	214-312-2206	bousquetd@aafes.com
Skolrick, Bobby	Army and Air Force Exchange Service 3911 S. Walton Walker Blvd. Dallas, TX 75236-1598	214-312-3148	214-312-2896	skolrick@aafes.com
Hampton, Karen L.	Army Records Management Division 6000 6 <sup>th</sup> Street - STOP C55 Fort Belvoir, VA 22060-5576	703-806-3720	703-806-3230	hamptonk@rmmpo.belvoir.army.mil
Prescott, Daryl R.	c3risk 7138 Little River Turnpike - Suite 900 Annandale, VA 22003	703-354-2682	703-354-0572	prescottldr@c3risk.com
Galyean, Richard	CACI, Inc. - Federal 1600 Wilson Blvd. - Suite 1300 Arlington, VA 22209-2515	703-558-0280	703-875-2904	rgalyean@hq.caci.com
Putman, Stephen J.	CACI, Inc. - Federal 1600 Wilson Blvd. - Suite 1300 Arlington, VA 22209-2515	703-558-0283	703-875-2904	sputnam@hq.caci.com
Lackey, Scott	Center for Army Lessons Learned 10 Meade Ave. Fort Leavenworth, KS 66027	913-684-5996	913-684-4387	lackeys@leav-emh1.army.mil

75

UNCLASSIFIED

GA22F042

UNCLASSIFIED

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Goodwin, Linda B.	Commandant of the Marine Corps 2 Navy Annex - Room 2128, ARSE Washington, DC 20380-1775	703-614-1081	703-693-7270	goodwinlb@hqmc.usmc.mil
Smith, Sherry D.	Comptroller of the Currency 250 E Street, SW - M/C 1-11 Washington, DC 20219	202-874-5035	202-874-5242	
Wohlhueter, Bob	Defense Contract Audit Agency (DCAA) 8725 John J. Kingman Road - Suite 2135 Fort Belvoir, VA 22060-6219	703-767-1036	703-767-1011	bob.wohlhueter@dcaa.mil
Bartolowits, Stephanie	Department of Defense ATTN: S. Bartolowits S541, Suite 6886 9800 Savage Road Ft. Meade, MD 20755-6886	301-688-0094	301-688-2342	
Cook, Susan A.	Department of Defense ATTN: S. Cook S541, Suite 6886 9800 Savage Road Ft. Meade, MD 20755-6886	301-688-0094	301-688-2342	
Lowery, Jackie E.	Department of Defense ATTN: J. Lowery S541, Suite 6886 9800 Savage Road Ft. Meade, MD 20755-6886	301-688-0094	301-688-2342	
Newlin, Burt	Department of Defense OSD (C3I) - Room 7012 Crystal Mall 3 Arlington, VA 22202	703-607-0244	703-607-0248	burt.newlin@osd.pentagon.mil
Smith, Doris V.	Department of Defense 14650 Monticello Drive Cocksville, MD 21723	301-688-0094	301-688-2065	

UNCLASSIFIED

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Wilt, Willard	Department of Defense ATTN: S. Cook S541, Suite 6886 9800 Savage Road Ft. Meade, MD 20755-6886	301-688-2272	301-688-2342	
Braemer, Raymond W.	Department of Defense Inspector General 400 Army Navy Drive - Suite 408 Arlington, VA 22202-2885	703-604-9781	703-604-9792 (call before sending FAX)	rbremer@dodig.osd.mil
Barth, Charley	Department of the Navy Chief Information Officer 1000 Navy Pentagon Washington, DC 20350-1000	703-602-6526	703-601-0352	barth.charley@hq.navy.mil
Raines, Erv	Department of the Navy Chief Information Officer 1000 Navy Pentagon Washington, DC 20350-1000	703-607-5652	703-601-0352	raines.erv@hq.navy.mil
Walter, Steven	Department of the Navy Chief Information Officer 1000 Navy Pentagon Washington, DC 20350-1000	703-607-5608	703-601-0352	walter.steve@hq.navy.mil
Johnson, Carolyn	Department of the Treasury U.S. Customs Service 1300 Pennsylvania Avenue, NW - Room 3.2C Washington, DC 20229	202-927-0744		
Williams, Laverne	Department of the Treasury U.S. Customs Service 1300 Pennsylvania Avenue, NW - Room 3.2CA Washington, DC 20229			
Matsuura, Stephen T.	DISA/JITC JITC/JTDB Fort Huachuca, AZ 85613	520-538-5169	520-538-4375	matsuurs@fhq.disa.mil
Jones, Jack	DTIC 8725 John J. Kingman Road - Suite 0944 Fort Belvoir, VA 22060-6218	703-767-8058	703-767-8065	jejones@dtic.mil

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Barry, Daniel A.	Executive Office of the President Office of Administration NEOB Room 4208 725 17 <sup>th</sup> Street, NW Washington, DC 20503	202-395-6476	202-395-3209	daniel_a_barry@oa.eop.gov
Doering, Nell	Executive Office of the President Office of Administration NEOB Room 4208 725 17 <sup>th</sup> Street, NW Washington, DC 20503	202-395-6471	202-395-7239	nellie_w_doering@oa.eop.gov
Miller, Dale F.	Executive Office of the President Office of Administration NEOB Room 4208 725 17 <sup>th</sup> Street, NW Washington, DC 20503	202-395-6470	202-395-7239	dale_f_miller@oa.eop.gov
Adler, Sanford C.	FBI Program Management Office Room 5853 935 Pennsylvania Avenue, NW Washington, DC 20535-0001	202-324-4850	202-324-5653	
Sinatra, Seung Ja	FDA/OC/OIRM Room 16B-26 - HFA-80 5600 Fishers Lane Rockville, MD 20857	301-827-4274	301-594-0060	ssinatra@oc.fda.gov
McGlasson, Lisa	Federal Aviation Administration 800 Independence Avenue, SW - APF-100 Washington, DC 20591	202-267-9885	202-267-5407	lisa.mcglasson@faa.gov
Daniels III, Phillip U.	Federal Bureau of Investigation 935 Pennsylvania Avenue, NW - Room 1376 Washington, DC	202-324-8781	202-324-5845	phillipdaniels@netscape.net
Deley, John Paul	Federal Trade Commission 6 <sup>th</sup> Street and Pennsylvania Avenue, NW - H-172 Washington, DC 20580	202-326-2085	202-326-2496	jdeley@ftc.gov

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Niner, Tom	ITT Systems DTRA ATTN: DASAC 6801 Telegraph Road Alexandria, VA 22310	703-325-8748	703-325-7002	tom.niner@hq.dswa.mil
Mahoney, Bette S.	Joint Requirements and Integration Office Defense Human Resources Activity 4015 Wilson Blvd. - Suite 1212 Arlington, VA 22203	703-696-8710	703-696-8703	mahoneyb@pr.osd.mil
Robson, Kyle	Labat-Anderson Inc. U.S. EPA, MC 3408 401 M Street, SW Washington, DC 20460	202-260-6969	202-401-5447	robson.kyle@epa.gov
Growth, Diane	Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041	703-845-8181	703-845-0127	diane.growth@lmco.com
Kivi, Judy	Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041	703-845-8181	703-845-0127	judy.f.kivi@lmco.com
Lampersberger, AnnMarie	Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041	703-845-8181	703-845-0127	annmarie.lampersberger@lmco.com
MacTavish, Sue	Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041	703-845-8181	703-845-0127	sue.h.mactavish@lmco.com
Pickard, Mike	Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041	703-845-8181	703-845-0127	michael.r.pickard@lmco.com
Schulke, Ed	Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041	703-845-8181	703-845-0127	edward.l.schulke@lmco.com
Wenchel, George	Lockheed Martin 5203 Leesburg Pike - Suite 1500 Falls Church, VA 22041	703-845-8181	703-845-0127	george.f.wenchel@lmco.com

UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Ambacher, Bruce	National Archives and Records Administration Center for Electronic Records 8601 Adelphi Road College Park, MD 20740	301-713-6630	301-713-6911	bruce.ambacher@arch2.nara.gov
Carlson, Michael	National Archives and Records Administration 8601 Adelphi Road College Park, MD 20740	301-713-7158		michael.carlson@arch2.nara.gov
Puglia, Steve	National Archives and Records Administration Special Media Preservation Branch 8601 Adelphi Road College Park, MD 20740	301-713-6713		steve.puglia@arch2.nara.gov
Thibodeau, Ken	National Archives and Records Administration Center for Electronic Records 8601 Adelphi Road College Park, MD 20740	301-713-6635	301-713-6911	ken.thibodeau@arch2.nara.gov
Conrad, Mark	National Historical Publications and Records Commission National Archives and Records Administration 7th Street and Pennsylvania Avenue, NW Washington, DC 20408			mark.conrad@arch1.nara.gov
Anderson, Russell G.	National Imagery & Mapping Agency ATTN: MSAR, N-42 1200 First Street, SE Washington, DC 20303-0001	202-314-1056	202-314-1099	andersonr@nima.mil
Watson, Leslie	National Institutes of Health 6011 Executive Boulevard - Room 601, MSC 7669 Rockville, MD 20892	301-496-2832	301-402-0169	LJ27G@nih.gov

80

UNCLASSIFIED

GA22F042

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Vogt-O'Connor, Diane	National Park Service 1849 C Street, NW - NC230 Washington, DC 20240	202-343-1011	202-343-1767	diane_vogt_o_connor@nps.gov
Campbell, Candi	National Reconnaissance Office (NRO) 14675 Lee Road Chantilly, VA 20151-1715	703-808-2474	703-808-5082	
Cumbra, John	National Reconnaissance Office (NRO) 14675 Lee Road Chantilly, VA 20151-1715	703-808-5500	703-808-5082	cumbrajo@nro.mil
Stawarz, Joanne	National Reconnaissance Office (NRO) 1399 Old Quincy Lane Reston, VA 20194	703-808-5016	703-808-5082	
Persons Jr., Henry W.	Naval Criminal Investigative Service WNY Bldg 111 Attn: Code 27D 716 Sicard Street, SE Washington, DC 20388-5380	202-433-9505	202-433-9518	hpersons@ncis.navy.mil
Rosser, Linda	NAVSEA	703-602-4875		
Shafer, Joan	NAVSEA	703-602-4875		
Slattery, Joy A.	NAVSEA, SEA OOI 303 Crossing Road Fredericksburg, VA 22406	703-602-4646, ext 317	703-602-2951	slatteryja@navsea.navy.mil
Parks, Hope L.	National Imagery & Mapping Agency 9414 Foxcroft Avenue Clinton, MD 20735	202-479-1923	202-314-1099	
Kelly, Anthony	Office of the Under Secretary of Defense for Policy 400 Army Navy Drive - Suite 100A Arlington, VA 22202	703-604-5523	703-604-4745	anthony.kelly@osd.pentagon.mil
Winter, Celia	Acting Director Information Management and Services Office of Thrift Supervision U.S. Treasury 1700 G Street, NW Washington, DC 20552	202-906-5909	202-906-7755	celia.winter@ots.treas.gov

UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Rawlings-Milton, Mary	Office of Thrift Supervision 1700 G Street, NW Washington, DC 20552	202-906-6028	202-906-7315	mary.rawlingsmilton@ots.treas.gov
Boyd, Deneise	Patent and Trademark Office Crystal Park 2, Room 310 2231 Crystal Drive Washington, DC 20231	703-305	703-305	deneise.boyd@uspto.gov
Fludd, Johnny	Patent and Trademark Office Crystal Park 2, Room 310 2231 Crystal Drive Washington, DC 20231	703-305	703-305	johnny.fludd@uspto.gov
Jones, Pola	Patent and Trademark Office Crystal Park 2, Room 310 2231 Crystal Drive Washington, DC 20231	703-305-9429	703-305	pola.jones@uspto.gov
Langan, Walter	Patent and Trademark Office Crystal Park 3, Room 310 2231 Crystal Drive Washington, DC 20231	703-305	703-305	walter.langan@uspto.gov
Meier, Linda	Patent and Trademark Office Crystal Park 3, Room 310 2231 Crystal Drive Washington, DC 20231	703-305	703-305	linda.meier@uspto.gov
Reid, Wendell G.	Office of Audit Railroad Retirement Board 844 North Rush Street - 4 <sup>th</sup> Floor Chicago, IL 60611	312-751-4954	312-751-4342	wgreid@yahoo.com
Lubbes, R. Kirk	Records Engineering, LLC 1567 Regatta Lane Reston, VA 20194	703-742-8558	703-742-8558	rec-engr@erols.com
Gowans, James	Smithsonian Institution Archives MRC 414/2135 Art 900 Jefferson Drive, SW Washington, DC 20560	202-357-1421, ext 43	202-357-2395	gowansj@ic.si.edu

82

UNCLASSIFIED

GA22F042

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Henson, Pamela M.	Smithsonian Institution Archives A & I 2135 Smithsonian Institution Washington, DC 20560-0414	202-786-2735	202-357-2395	hensonp@osia.si.edu
Nichols, Jennifer	Smithsonian Institution Archives Institutional History Division A & I 2135 Smithsonian Institution Washington, DC 20560-0414	202-357-1421, ext 41	202-357-2395	nicholsj@osia.si.edu
Petty, Jeremy	TechLaw Inc. 750 N. St. Paul - Suite #600 Dallas, TX 75201	214-953-0045		jpetty@techlawinc.com
Chernikoff, Tod	The Marasco Newton Group, Ltd. 2801 Clarendon Boulevard Arlington, VA 22201	703-292-5844	703-413-0222	tchernik@marasconewton.com OR chernikoff.tod@epa.gov
Senol, Annie	The Marasco Newton Group, Ltd. 2801 Clarendon Boulevard Arlington, VA 22201	703-292-5843	703-413-0222	asenol@marasconewton.com
Chadduck, Robert	National Archives and Records Administration 8601 Adelphi Road College Park, MD 20740	301-713-6640, ext 301	301-713-6911	robert.chadduck@arch2.nara.com
Worthington, Linda K.	U.S. Army Corps of Engineers 20 Massachusetts Avenue, NW Washington, DC 20314-1000	202-761-0332		
Wehrly, Stephen P.	U.S. Army Publishing Agency Publishing Division Chief, Electronic Publishing Branch JDHQSVP-PAP-E, Room 1002 2461 Eisenhower Avenue Alexandria, VA 22331-0302	703-325-6218	703-325-6260	wehrlys@usapa.army.mil
Hix, Matthew	U.S. Department of Education 400 Maryland Avenue, SW - ROB-3, Room 5624 Washington, DC 20202-5641	202-708-9265	202-708-9346	matthew_hix@ed.gov

## UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

NAME	ORGANIZATION AND MAILING ADDRESS	TELEPHONE NUMBER	FAX NUMBER	E-MAIL ADDRESS
Thomas, Chiquitta	U.S. Department of Education 400 Maryland Avenue, SW - ROB-3, Room 5624 Washington, DC 20202-5641	202-708-9265	202-708-9346	chiquitta_thomas@ed.gov
McFarland, Verne	U.S. Environmental Protection Agency	214-665-6617	214-665-6660	mcfarland.verne@epamail.epa.gov
McHale, Laura	U.S. Environmental Protection Agency Mail Code 3408 401 M Street, SW Washington, DC 20460	202-260-9329	202-401-8390	mchale.laura@epa.gov
Goewey, Robert J.	U.S. Marshals Service 600 Army Navy Drive Arlington, VA 22202	202-302-5187	202-307-8729	bob.goewey@usdoj.gov
Curtis, Sharon	USDA	202-720-9270	202-205-2831	sharon.curtis@usda.gov
Richardson, Kimberly	USDA 5601 Sunnyside Avenue - Mail Drop 5242 Beltsville, MD 20705-5242	301-504-4233		kimberly.richardson@usda.gov
McLendon, Vern	USDA - FSIS 5601 Sunnyside Avenue - Mail Drop 5242 Beltsville, MD 20705-5242	301-504-4215	301-504-4281	vern.mclendon@usda.gov
Rardon, Sherri D.	USDA - FSIS 5601 Sunnyside Avenue - Mail Drop 5242 Beltsville, MD 20705-5242	301-504-4215	301-504-4281	sherri.rardon@usda.gov
Calhoun, Barbara	USDB - DIA 300 McPherson Avenue Fort Leavenworth, KS 66027-1363	913-684-4629	913-684-4360	calhoub@leav-emb1.army.mil
Swenson, Suzanne	Zimmerman Associates Inc. 8229 Boone Blvd. - Suite 300 Vienna, VA 22182	703-883-0506	703-883-0526	sswenson@zai-amelex.com

UNCLASSIFIED

GA22F042

UNCLASSIFIED

1 June 1999

(This space intentionally left blank.)

UNCLASSIFIED

GA22F042

## APPENDIX G - ACRONYMS

AACR2	Anglo-American Cataloguing Rules, Second Edition
AFNOR	Association Francaise de Normalisation
AIIM	Association for Information and Image Management
ANSI	American National Standards Institute
ANSI Z39.50	Information Retrieval Application Service Definition and Protocol Specification
ASCII	American Standard Code for Information Interchange
AWG	Automation Working Group, a part of DPMC
BIIF	Basic Image Interchange Format
BMP	Windows Bitmap
CAD	Computer Aided Design
CALS	Continuous Acquisition and Life-Cycle Support
CCITT	International Telegraph and Telephone Consultative Committee – Facsimile Compression group 4
CD	Compact Disk
CDR	Corel Draw Format
CD-ROM	Computer Disc – Read Only Memory
CFF2	Common File Format, Revision 2
CFR	Code of Federal Regulation
CGM	Computer Graphics Metafile
CGRM	Computer Graphics Reference Model
COM	Computer-Output Microfilm
COTS	Commercial-Off-The-Shelf
CUT	Media Cybernetic's Dr. Halo Graphic Format
DAPS	Defense Automated Printing Service
DCT	Discrete Cosine Transform
DDES	Digital Data Exchange Specification
DIS	Document Interchange System
DISA	Defense Information Services Agency
DMS	Document Management System
DoD	Department of Defense
dpi	Dots Per Inch
DPMC	Declassification Program Managers Council
DPRC	Declassification Productivity Research Center
DPRIS	Defense Personnel Records Imaging System
DRW	Micrografix Designer Format
DTD	Document Type Definition
DTP	Desktop Publishing

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

DWG	An AutoCAD two-dimensional drawing file format
DXF	Data Exchange Format; Drawing Interchange Format
EAD	Encoded Archival Description
EBCDIC	Extended Binary Coded Decimal Interchange Code
ECMA	European Computer Manufacturers Association
EDIS	Electronic Document Interchange Standard
ETM	Electronic Technical Manual
FDIS	Final Draft International Standard
FGDC	Federal Geospatial Data Committee
GEF	Graphics Exchange Format
GEM	Digital Research's GEM Metafile Format
GGCA	Geometric Graphics Content Architecture
GIF	Graphics Interchange Format
GILS	Government Information Locator Service
GKS	Graphical Kernel System
GKS-3D	Graphical Kernel System - 3 Dimensions
HPGL	Hewlett Packard Graphics Language
HTML	Hypertext Mark-up Language
HW	Hardware
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
IGES	Initial Graphic Exchange Specification
IIF	Image Interchange Format
ILBM	Interleaved Bitmap
IMG	GEM IMG
IMJ	Image JPEG
IPI	Image Processing and Interchange
IPI-IIF	Image Processing and Interchange: Image Interchange Facility
ISO	International Organization for Standardization
ISP	International Standardized Profiles
IT	Information Technology
ITU	International Telecommunication Union
JBIG	Joint Bi-Level Imaging Group
JFIF	JPEG File Interchange Format
JPEG	Joint Photographic Experts Group
JTC1	Joint Technical Committee 1 of the ISO/IEC

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

LC	Library of Congress
LOGSA	Logistics Support Activity
LZW	Lempel Ziff Welch; Lempel, Ziv and Welch
MARC	Machine Readable Cataloging
MPR	Military Personnel Records
MSP	Microsoft Paint
NARA	National Archives and Records Administration
NATO	North Atlantic Treaty Organization
NBS	National Bureau of Standards
NIMA	National Imagery and Mapping Agency
NIS	National Institute of Standards
NITFS	National Imagery Transmission Format Standard
NPRC	National Personnel Records Center
OASD/C3I	Office of the Assistant Secretary of Defense (Command, Control, Communications and Intelligence)
OCR	Optical Character Recognition
ODA	Open Document Architecture and Interchange Format; Open Document Architecture
ODA GGCA	ODA Geometric Graphics Content Architecture
ODA RGCA	ODA Raster Graphics Content Architecture
OMPF	Official Military Personnel Files
OII	Open Information Interchange (European Commission)
OPR	Organization of Primary Responsibility
OSI	Organization of Secondary Interest
PBM	Portable Bit Map
PCX	PC Paintbrush
PDF	Portable Document Format
PHIGS	Programmer's Hierarchical Interactive Graphics System
PIC	Lotus 1-2-3 Graphic Interchange File
PICT	Macintosh Picture; Apple's Picture Format
PIKS	Programmer's Imaging Kernel System
PNG	Portable Network Graphics
PNTG	Apple's MacPaint Format
RDF	Resource Description Framework
RGB	Red, Green, Blue
RLE	Run length encoded
RLG	Research Libraries Group
RMS	Records Management System

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

SCR	Microsoft's Screen Capture Format
SET	Secure Electronic Transactions; Standard d'Echange et de Transfert
SGML	Standard Generalized Mark-up Language
SPIFF	Still Picture Interchange File Format
STEP	Standard for the Exchange of Product Model Data; Product Data Representation and Exchange
SVG	Scalable Vector Graphics
SW	Software
TEI	Text Encoding Initiative
TGA	TARGA - 24 bit true color
TIFF	Tag Image File Format
TIFF/IT	TIFF for Image Technology
UPF	Universal Preservation Format
USC	United States Code
USAPA	U.S. Army Publishing Agency
US MARC	United States version of Machine Readable Cataloguing
VDAFS	VDA Surface Interface
VRML	Virtual Reality Modeling Language
W3C	World Wide Web Consortium
WMF	Windows metaformat - raster only; Microsoft's Windows Metafile Format
WORM	Write Once-Read Many
WPG	WordPerfect graphic format - raster only
WWW	World Wide Web
XBM	X Windows Bitmap
XML	eXtensible Mark-up Language
XWD	X Windows Dump

## APPENDIX H - DEFINITIONS

ABIC	IBM image compression for check scanners. Interface to IBM library.
Agency	Any executive department or independent establishment in the executive branch of the Government, including any wholly-owned Government corporation (36 CFR 1220.14)
Alpha Channel	Seamless image integration with transparency.
Anti-Aliasing	A method of representing data which has been missed due to under-sampling or when an image is reduced in resolution (for example, when a 300 dpi image is converted to 96 dpi for display). One of the most common benefits is preserving lines or complete characters which would otherwise appear broken or disappear.
API	Application Programmer's Interface. The command set for a set of routines that invoke a library or toolkit component.
Aspect Ratio	The proportion of an image's size given in terms of the horizontal dimension versus the vertical dimension. An aspect ratio of 4:3 indicates that the image is 4/3 times as wide as it is high.
AVI	Rasterized video format designed to allow moving pictures to be stored and played back on computers.
Bitmap	An image is called a bitmap or raster image if its objects or contents are represented by pixels. This is the opposite of a vector representation image where objects are described by beginning and endpoints for lines, and center and radius for circles and ellipses.
Bitonal Image	An image comprised of pixels that contain only a single bit of information. Each pixel is either on or off. Normally, "on" is white and "off" is black. FAX image formats and Group 4 image formats are bitonal images.
Call-back Function	A call-back function is a function that is passed to another function as a parameter. The function receiving the call-back function can then call this function. This is a powerful programming method used to change the behavior of a given routine.
CMYK	Cyan, Magenta, Yellow, (K) black. The four planes of color used in the pre-press industry to represent images to be printed.
Compression	A process of encoding image or other data so that it occupies less memory or disk space than its uncompressed version.
Coordinate Review	A declassification review and/or release by two or more Agencies having an equity interest in a document. Sometimes also called a "coordinated review," "external review," or "equity review."
Crop	An image processing method of selecting a rectangular region of the image for removal.
DDB	Device Dependent Bitmap. A bitmap dependent upon a particular hardware device.
Decompression	The method or process of restoring a compressed image or file to its original form.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

DIB	Device Independent Bitmap is an image format specification independent of all hardware devices and platforms.
Digital Image Formats	The digital image format selected for this standard is TIFF-6. This is the standard image format for most image-scanning devices, and is probably the most widely supported standard. TIFF permits the insertion of user-defined information into the header of the image file by means of "tags". The proposed implementation of the standard has been carefully selected to permit agencies to use other formats if desired; the only requirement is that the software that views the images be OLE 2.0 compatible.
Dithering	A method of using similarly colored and sized pixels to display or print a different color or resolution.
DLL	Dynamic Linked Library. A compiled and linked collection of computer functions that are not immediately bound to an executable (EXE) but are called during program execution.
Document	The same as NARA's used of the term "records," namely: all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations or other activities of the Government or because of the informational value of the data in them (44 U.S.C. 3301).
DPI	Dots Per Inch. A measure of the resolution of electronic images; the higher the number, the more fidelity the electronic image has to the original document appearance.
E.O 12958	Executive Order 12958 defines "national security information" and requirements for classification and declassification. President Clinton issued it on April 20, 1995. It states among other things that all classified records of permanent historical interest more than 25 years old shall be automatically declassified in April 2000, unless the Executive Branch Agencies having equities in the documents can give a reason for exemption from declassification. Nine exemption categories are specified. It is currently estimated that there are over 700 million pages of classified permanent records material that is 25 years old or older subject to automatic declassification review.
Eight Bit Image	An image where each pixel has 8-bits of information in it. An 8-bit pixel can take on one of 256 possible values. There are two common types of 8-bit images: gray scale and palette. In gray-scale images each pixel takes on one of 256 shades of gray and the shades are linearly distributed from 0 (black) to 256 (white). For 8-bit color, each pixel is used as an index into the palette. Thus these images can have up to 256 different colors in them at one time. Indexed 8-bit images are good for low color resolution images.
EPS	Writes full Postscript. Reads any embedded raster.
Exemptions	E.O. 12958 and FOIA specify a number of conditions that may exempt documents from declassification and/or release.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

External Referral Processing Information	This includes data about the transmittal of media among Agencies. It may include procedures, identifiers and courier information which is clearly specified in Government standards.
File Format	A specification for storing image data. The format dictates what information is present in the file and how it is organized within it.
Flashpix	Kodak, Hewlett Packard, Microsoft Color Imaging format – read only.
FOIA	The Freedom of Information Act specifies how individuals may request information from agencies and requires that the agencies review and attempt to declassify the information so requested, meaning that documents must be reviewed line-by-line and redacted as necessary.
Grayscale Images	Images of each page of an original grayscale document shall be passed using 8-bits/pixel of tonal depth. This allows 256 grayscale levels.
Group 3	CCITT Group 3 (1D and 2D) – bitonal, used for FAX
Group 4	CCITT Group 4 – bitonal, used for document imaging
GUI	Graphical User Interface. A computer interface which uses graphical objects.
Half-tone	The reproduction of a continuous-tone image on a device which does not directly support continuous output. This is done by displaying or printing pattern of small dots which from a distance can simulate the desired output color or intensity.
Image	An image of a document is the electronic version of a pre-existing physical document. Images may be created by digital camera or by electronic scanner. An image can also be the original electronic version of the document, as long as the document's native format was used for the specific purpose of creating an electronic image that would be maintained as the master document. An example is a Federal agency's policy manual that was generated in either HTML or SGML for the purpose of making the document available on the World Wide Web (WWW).
Image Compression	For bitonal images, the Standard's compression method shall be CCITT Group 4 (lossless) compression, as implemented in TIFF 6.0. For color and grayscale images, the Standard's compression method shall be CCITT Group 4 (lossless) compression. If that is not in fact available when needed, then JPEG (lossless) compression or other compression method agreed to by bilateral agreement shall be used.
Image Depth	Images can be scanned and stored at a wide range of depths, from 2 colors (bitonal) to 16 colors (grayscale), 256 colors (8-bits), 65,536 colors (16-bits), or 16,777,216 colors (24-bits). The standard supports a variety of image depths, dependent upon the original document and OPR internal requirements.
Image Resolution	Scanning at 300 DPI is currently widely accepted for electronic document management purposes. To avoid confusion when digitizing from an intermediate copy of a record (e.g. microfilm), the intention of the Standard is to scan the documents original size at 300 DPI, not its size as reduced or enlarged.
JBIG	Proprietary bitonal compression format (interface to IBM library).

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

JPEG	Joint Photographic Experts Group, a set of 29 digital image coding processes developed by computer graphics organizations for achieving both high compression and high fidelity of images. It is an encoding format, not an actual file format.
Konica	Konica color format.
Lempel Ziff Welch (LZW)	An image compression method found in the popular GIF format and patented by Unisys.
Library	A collection of software functions that can be called upon by a higher level program. Most libraries are collections of similar routines such as those used for graphical or image processing.
Lossless	A method of image compression where there is no loss in quality when the image is compressed or uncompressed.
Lossy	A method of image compression where some image quality is sacrificed in exchange for higher compression ratios. The most common lossy image compression method is JPEG.
Mel0	Two-dimensional CAD product from Hewlett-Packard
Metadata	Metadata is information about a document, such as its identification, author, title, current classification, etc. Metadata can be divided into two broad categories: data for which only one instance is expected to occur (e.g., document ID, apparent classification) and data for which an unknown number of instances could occur (e.g., authors, recipients, reviewers, dates, exemption codes). This latter type of data is called repeating metadata. Perhaps one of the most important decisions to be made in the selection of an imaging standard(s) is the trade-off between image file size (therefore cost) and the need to retain detail for future users (including historians and the public).
MPEG	Motion Pictures Experts Group. An ISO specification of the compression of digital-broadcast quality full-motion video with its sound track.
MPEG-1	Rasterized video format designed to allow moving pictures to be stored and played back on computers.
MPEG-2	Rasterized video format designed to allow moving pictures to be stored and played back on computers.
Number of Page Images per File	Generally, under this Standard, all pages of a document shall be stored in a single TIFF file. This will facilitate the use of software to automatically collect and combine redaction data from the headers of multiple TIFF files (presumably coming from different agencies or subgroups within an agency). This will greatly improve final review productivity, compared to manually attempting to consolidate redactions proposed by multiple sources. If using a single page image per file TIFF format, a written bilateral agreement is recommended to ensure compatibility among Agency pairs. The proposed reference implementation can technically support either approach; the only requirement is that the software that views the images be OLE 2.0 compatible.
OCR	Optical Character Recognition. A process for reading scanned document images and producing corresponding ASCII text.

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

OPR	Organization of Primary Responsibility. Any Agency responsible for the declassification review of a particular document.
OSI	Organization of Secondary Interest. Any Agency an OPR believes has an equity in a particular document and should conduct a declassification review.
Palette	A digital image palette is a collection of 3 look-up-tables which are used to define a given pixel's display color. One table for red, one for green, and one for blue.
Pixel	A digital image is made up of rows and columns of points. Each point is called a pixel. Each pixel in an image is addressed by its column (x) and its row (y). An 8-bit pixel can take on one of 256 values. A 24-bit pixel image usually has three 8-bit components for each of the primary colors: red, green, and blue.
PNG	8 and 24-bit raster format, replacement for GIF and LZW (read and write).
Quicktime	Rasterized video format designed to allow moving pictures to be stored and played back on computers.
Raster	A term, which for historical reasons, is used to describe a single row of a digital image. Thus, a raster image is one that is made up of rows of pixels.
Records	All books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations or other activities of the Government or because of the informational value of the data in them. (44 USC 3301)
Resolution	Image resolution is the number of pixels per unit of length along the x and y axis.
RGB	Red, Green, Blue. A triplet of numeric values which are used to describe a color.
Screen Coordinates	Screen coordinates are those of the actual graphics display controller. The origin is almost always at the upper left-hand corner of the display.
Standard	<p><i>De jure</i> – A publicly available definition of a hardware or software component, resulting from international, national, or industrial agreement. For example BSI (British Standards Institute) or ISO.</p> <p><i>De facto</i> – When certain formats and designs acquire a sufficient market position to be accepted without legal validation. i.e. no standard agreement has been formulated. For example, Microsoft Windows.</p> <p><i>Industry/vendor based</i> – The development and evolution of a standard by an industrial/vendor based group rather than by a formal standards committee.</p>
Thumbnail	A small, typically low resolution representation of an image. Usually used to display many images on the screen at once.
TIFF Header Tags	The only document metadata to be stored in the TIFF header shall be: document ID, which shall be stored in the standard TIFF "DocumentName" tag (Tag #10DH), temporary annotations (if any), including color-coded overlays identifying areas to be redacted, and their associated exemption codes (Tag #32932). While annotations shall

# UNCLASSIFIED

Imaging Standard Support Task

1 June 1999

be stored in the TIFF header for exchange purposes, each agency should ensure that any final release version of images of redacted documents to be released shall have no metadata in the TIFF header except the ESDN.

Tonal (Color) Images	Images of each page of an original colored document shall be passed using 24-bits/pixel of tonal depth. This allows 16,777,216 colors.
Twenty-Four (24)-bit Image	A 24-bit image contains pixels which are made up of RGB triplets.
Two-Tone (Black and White) Images	Images of each page of an original two-tone document shall be passed using 1-bit/pixel of tonal depth. This allows 2 grayscale levels, also called bitonal or black-and-white imagery.
Wavelet compression	Replacement for JPEG compression.
Winfax	FAX format